

Db 181 TWICTVLQNKQKVEFKIDIV 200

RESULT 28

US-08-328-500-9

Sequence 9, Application US/08328500

Patent No. 6673896

GENERAL INFORMATION:

APPLICANT: Maddon, Paul J.

APPLICANT: Axel, Richard

APPLICANT: Sweet, Richard W.

APPLICANT: Arthos, James

TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/328,500

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 0575/24577-CY

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 457 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-328-500-9

Query Match 78.1%; Score 1029; DB 4; Length 457;
Best Local Similarity 100.0%; Pred. No. 3,4e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLLVQLALLPAATQGNKRVLGKGDVLTCTASQKKSIOFHWNKSNQIK 60
Db 1 MNRGVPRHLLLVQLALLPAATQGNKRVLGKGDVLTCTASQKKSIOFHWNKSNQIK 60
QY 61 ILGNQGSFLTKGPKSNDRADSRSLMDQGNPFLIKNLKIEDSDTYICEVEDQKEEVOL 120
Db 61 ILGNQGSFLTKGPKSNDRADSRSLMDQGNPFLIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSPTLHLLQGSGLTLTLSPSPSSPVQCRSPRGKNIQGGKTLISVQLELQDSG 180
Db 121 LVFGLTANSPTLHLLQGSGLTLTLSPSPSSPVQCRSPRGKNIQGGKTLISVQLELQDSG 180
QY 181 TWICTVLQNKQKVEFKIDIV 200
Db 181 TWICTVLQNKQKVEFKIDIV 200

RESULT 29

US-08-466-368-4

Sequence 4, Application US/08466368

Patent No. 6093539

GENERAL INFORMATION:

APPLICANT: Maddon, Paul J.

APPLICANT: Littman, Dan R.

APPLICANT: Chess, Leonard

APPLICANT: Axel, Richard

APPLICANT: Weiss, Robin

APPLICANT: McDougall, J. S.

TITLE OF INVENTION: DNA ENCODING THE T CELL SURFACE PROTEIN

TITLE OF INVENTION: T4 AND USE OF FRAGMENTS OF T4 IN THE TREATMENT OF AIDS

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP

STREET: 1185 Avenue of Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,368

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 24577-BI-B/JPM/AKC

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-278-0400

TELEFAX: 212-391-0525

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 458 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-466-368-4

Query Match 78.1%; Score 1029; DB 3; Length 458;
Best Local Similarity 100.0%; Pred. No. 3,4e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLLVQLALLPAATQGNKRVLGKGDVLTCTASQKKSIOFHWNKSNQIK 60
Db 1 MNRGVPRHLLLVQLALLPAATQGNKRVLGKGDVLTCTASQKKSIOFHWNKSNQIK 60
QY 61 ILGNQGSFLTKGPKSNDRADSRSLMDQGNPFLIKNLKIEDSDTYICEVEDQKEEVOL 120
Db 61 ILGNQGSFLTKGPKSNDRADSRSLMDQGNPFLIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSPTLHLLQGSGLTLTLSPSPSSPVQCRSPRGKNIQGGKTLISVQLELQDSG 180
Db 121 LVFGLTANSPTLHLLQGSGLTLTLSPSPSSPVQCRSPRGKNIQGGKTLISVQLELQDSG 180
QY 181 TWICTVLQNKQKVEFKIDIV 200
Db 181 TWICTVLQNKQKVEFKIDIV 200

RESULT 30

US-08-417-495-5

Sequence 5, Application US/08417495

Patent No. 5843728

GENERAL INFORMATION:

APPLICANT: Seed, Brian et al.

TITLE OF INVENTION: Redirection of Cellular Immunity by

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

Chimeras

```
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/417,495
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/203,866
FILING DATE:
APPLICATION NUMBER: US/07/847,566
FILING DATE:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-417-495-5
```

```
Query March 78.1% Score 1029; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 3.4e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MNRGVPRHLLVQLALPPATQGNKVVLGKKGDTVELTCTASQKKSIOFHMKNNSNOIK 60
DB 1 MNRGVPRHLLVQLALPPATQGNKVVLGKKGDTVELTCTASQKKSIOFHMKNNSNOIK 60
QY 61 ILGNQGSFLLTKGPKSKLNDRADSRSLMDQGNFPLIKNLKIEBDDTYICEVEDQKEVOL 120
DB 61 ILGNQGSFLLTKGPKSKLNDRADSRSLMDQGNFPLIKNLKIEBDDTYICEVEDQKEVOL 120
QY 121 LVFGLTANSDTHLLQGSQSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLQGSQSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNKQKVEFKIDIV 200
DB 181 TWTCTVLQNKQKVEFKIDIV 200
```

```
RESULT 31
US-08-284-391B-5
Sequence 5, Application US/08284391B
Patent No. 5851828
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Ebling LLP
STREET: 176 Federal Street
CITY: Boston
```

```
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391B
FILING DATE: 02-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Ebling, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-284-391B-5
```

```
Query March 78.1% Score 1029; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 3.4e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MNRGVPRHLLVQLALPPATQGNKVVLGKKGDTVELTCTASQKKSIOFHMKNNSNOIK 60
DB 1 MNRGVPRHLLVQLALPPATQGNKVVLGKKGDTVELTCTASQKKSIOFHMKNNSNOIK 60
QY 61 ILGNQGSFLLTKGPKSKLNDRADSRSLMDQGNFPLIKNLKIEBDDTYICEVEDQKEVOL 120
DB 61 ILGNQGSFLLTKGPKSKLNDRADSRSLMDQGNFPLIKNLKIEBDDTYICEVEDQKEVOL 120
QY 121 LVFGLTANSDTHLLQGSQSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLQGSQSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNKQKVEFKIDIV 200
DB 181 TWTCTVLQNKQKVEFKIDIV 200
```

```
RESULT 32
US-09-218-950-5
Sequence 5, Application US/09218950
Patent No. 6284240
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Ebling LLP
STREET: 176 Federal Street
CITY: Boston
```

```
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Pasteo for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/218.950
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/284.391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195.395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847.566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665.961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L
REGISTRATION NUMBER: 35.238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-218-950-5

Query Match      78.1%; Score 1029; DB 3; Length 462;
Best Local Similarity 100.0%; Pred. No. 3.4e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLILVQLALPAATQGNKVVLGKGDVLELTCTASQKSIQPHMKNNOIK 60
DB 1 MNRGVPFRHLILVQLALPAATQGNKVVLGKGDVLELTCTASQKSIQPHMKNNOIK 60
QY 61 ILGNQGSFLTKGPKSKLNDRAISRSLMDQGNFPLIKLKIETSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPKSKLNDRAISRSLMDQGNFPLIKLKIETSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNKQKVEFKIDIV 200
DB 181 TWTCTVLQNKQKVEFKIDIV 200

RESULT 33
PCT-US92-01785-5
; Sequence 5, Application PC/TUS9201785
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation
; TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor
; TITLE OF INVENTION: Chimerae
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
```

```
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/01785
FILING DATE: 19920306
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665.961
FILING DATE: March 7, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-01785-5

Query Match      78.1%; Score 1029; DB 5; Length 462;
Best Local Similarity 100.0%; Pred. No. 3.4e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLILVQLALPAATQGNKVVLGKGDVLELTCTASQKSIQPHMKNNOIK 60
DB 1 MNRGVPFRHLILVQLALPAATQGNKVVLGKGDVLELTCTASQKSIQPHMKNNOIK 60
QY 61 ILGNQGSFLTKGPKSKLNDRAISRSLMDQGNFPLIKLKIETSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPKSKLNDRAISRSLMDQGNFPLIKLKIETSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNKQKVEFKIDIV 200
DB 181 TWTCTVLQNKQKVEFKIDIV 200

RESULT 34
PCT-US95-00454-5
; Sequence 5, Application PC/TUS9500454
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian et al.
; TITLE OF INVENTION: Targeted Cytolysis of HIV-Infected
; TITLE OF INVENTION: Cells by Chimeric CD4 Receptor-
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00454
```

FILED DATE: 07/847,566
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION: NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/247001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-00454-5

Query Match 78.1%; Score 1029; DB 5; Length 462;
Best Local Similarity 100.0%; Pred. No. 3,4e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVYLQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMNSNOIK 60
DB 1 MNRGVPRHLLVYLQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMNSNOIK 60
QY 61 ILGNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIKNLKIETSDYICEVEDQKEVOL 120
DB 61 ILGNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIKNLKIETSDYICEVEDQKEVOL 120
QY 121 LVFGLTANSPTHLIQGOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLSSVQLDQSG 180
DB 121 LVFGLTANSPTHLIQGOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLSSVQLDQSG 180
QY 181 TWTCTVLOQOKKVEFKIDIV 200
DB 181 TWTCTVLOQOKKVEFKIDIV 200

RESULT 35
US-08-417-495-6
Sequence 6, Application US/08417495
Patent No. 5843728
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Redirection of Cellular Immunity by Chimera
NUMBER OF INVENTIONS: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA: US/08/417,495
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/203,866
FILING DATE: 07/847,566
APPLICATION NUMBER: US/07/847,566
FILING DATE: March 6, 1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION: NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-417-495-6

Query Match 78.1%; Score 1029; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 4,1e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVYLQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMNSNOIK 60
DB 1 MNRGVPRHLLVYLQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMNSNOIK 60
QY 61 ILGNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIKNLKIETSDYICEVEDQKEVOL 120
DB 61 ILGNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIKNLKIETSDYICEVEDQKEVOL 120
QY 121 LVFGLTANSPTHLIQGOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLSSVQLDQSG 180
DB 121 LVFGLTANSPTHLIQGOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLSSVQLDQSG 180
QY 181 TWTCTVLOQOKKVEFKIDIV 200
DB 181 TWTCTVLOQOKKVEFKIDIV 200

RESULT 36
US-08-284-391B-6
Sequence 6, Application US/08284391B
Patent No. 5851828
GENERAL INFORMATION:
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF INVENTIONS: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elding LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA: US/08/284,391B
FILING DATE: 02-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Ebling, Karen L
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-284-391B-6

Query Match 78.1%; Score 1029; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 4.1e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVQLALPAAATQGNKVLGKGGDTVELTCTASOKKSIOFHMNSQIK 60
DB 1 MNRGVPFRHLVQLALPAAATQGNKVLGKGGDTVELTCTASOKKSIOFHMNSQIK 60
QY 61 ILGNQGSFLTKGPKSLNDRADSRSLWDQGNFPLIINKLIEDSDTYICEVEDQKEVQL 120
DB 61 ILGNQGSFLTKGPKSLNDRADSRSLWDQGNFPLIINKLIEDSDTYICEVEDQKEVQL 120
QY 121 LVFGLTANSPTHLQOGSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLISVSQLELDG 180
DB 121 LVFGLTANSPTHLQOGSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLISVSQLELDG 180
QY 181 TWICTVLQONQKVEFKIDIV 200
DB 181 TWICTVLQONQKVEFKIDIV 200

RESULT 37
US-09-218-950-6
Sequence 6, Application US/09218950
Patent No. 6284240
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Ebling LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PASCSEO for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/218,950
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/284,391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Ebling, Karen L
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-218-950-6

Query Match 78.1%; Score 1029; DB 3; Length 532;
Best Local Similarity 100.0%; Pred. No. 4.1e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVQLALPAAATQGNKVLGKGGDTVELTCTASOKKSIOFHMNSQIK 60
DB 1 MNRGVPFRHLVQLALPAAATQGNKVLGKGGDTVELTCTASOKKSIOFHMNSQIK 60
QY 61 ILGNQGSFLTKGPKSLNDRADSRSLWDQGNFPLIINKLIEDSDTYICEVEDQKEVQL 120
DB 61 ILGNQGSFLTKGPKSLNDRADSRSLWDQGNFPLIINKLIEDSDTYICEVEDQKEVQL 120
QY 121 LVFGLTANSPTHLQOGSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLISVSQLELDG 180
DB 121 LVFGLTANSPTHLQOGSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLISVSQLELDG 180
QY 181 TWICTVLQONQKVEFKIDIV 200
DB 181 TWICTVLQONQKVEFKIDIV 200

RESULT 38
PCT-US92-01785-6
Sequence 6, Application PC/TUS9201785
GENERAL INFORMATION:
APPLICANT: The General Hospital Corporation
TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/01785
FILING DATE: 19920306
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991

ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-01785-6

Query Match 78.1%; Score 1029; DB 5; Length 532;

Best Local Similarity 100.0%; Pred. No. 4,1e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLQALPPATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNSQIK 60
DB 1 MNRGVPRHLLVLQALPPATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNSQIK 60
QY 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120
DB 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120
QY 121 LVFGLTANSDTHLQGSLLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLQGSLLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNKVKVEFKIDIV 200
DB 181 TWTCTVLQNKVKVEFKIDIV 200

RESULT 39

PCT-US95-00454-6
Sequence 6, Application PC/TUS9500454
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Targeted Cycolysis of HIV-Infected
TITLE OF INVENTION: Cells by Chimeric CD4 Receptor-
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00454
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-00454-6

Query Match 78.1%; Score 1029; DB 5; Length 532;

Best Local Similarity 100.0%; Pred. No. 4,1e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLQALPPATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNSQIK 60
DB 1 MNRGVPRHLLVLQALPPATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNSQIK 60
QY 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120
DB 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120
QY 121 LVFGLTANSDTHLQGSLLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLQGSLLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNKVKVEFKIDIV 200
DB 181 TWTCTVLQNKVKVEFKIDIV 200

RESULT 40

US-08-417-495-4
Sequence 4, Application US/08417495
Patent No. 5843728
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Redirection of Cellular Immunity by
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/417,495
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/203,866
FILING DATE:
APPLICATION NUMBER: US/07/847,566
FILING DATE:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-417-495-4

Query Match 78.1%; Score 1029; DB 2; Length 575;
Best Local Similarity 100.0%; Pred. No. 4.6e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRVPRHLLVQLALPPAATGKRVVGGKDTVELCTASQKSIQFHWNQSNQIK 60
DB 1 MNRVPRHLLVQLALPPAATGKRVVGGKDTVELCTASQKSIQFHWNQSNQIK 60
QY 61 ILGQGSFLTKGPKSLNDRADSRSLMDQGNFPLIKLIEDSDTYICEVEDQKEEVQL 120
DB 61 ILGQGSFLTKGPKSLNDRADSRSLMDQGNFPLIKLIEDSDTYICEVEDQKEEVQL 120
QY 121 LVFGLTANSDTHLLQGSLTTLTSPGSSPSVQCRSPRGKNIQGKTLVSQLELDPSG 180
DB 121 LVFGLTANSDTHLLQGSLTTLTSPGSSPSVQCRSPRGKNIQGKTLVSQLELDPSG 180
QY 181 TWCTVQLONOKVEFKIDIV 200
DB 181 TWCTVQLONOKVEFKIDIV 200

RESULT 41

US-08-284-391B-4
Sequence 4, Application US/08284391B
Patent No. 5851828

GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391B
FILING DATE: 02-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991

ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045

TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-284-391B-4

Query Match 78.1%; Score 1029; DB 2; Length 575;
Best Local Similarity 100.0%; Pred. No. 4.6e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRVPRHLLVQLALPPAATGKRVVGGKDTVELCTASQKSIQFHWNQSNQIK 60
DB 1 MNRVPRHLLVQLALPPAATGKRVVGGKDTVELCTASQKSIQFHWNQSNQIK 60
QY 61 ILGQGSFLTKGPKSLNDRADSRSLMDQGNFPLIKLIEDSDTYICEVEDQKEEVQL 120
DB 61 ILGQGSFLTKGPKSLNDRADSRSLMDQGNFPLIKLIEDSDTYICEVEDQKEEVQL 120
QY 121 LVFGLTANSDTHLLQGSLTTLTSPGSSPSVQCRSPRGKNIQGKTLVSQLELDPSG 180
DB 121 LVFGLTANSDTHLLQGSLTTLTSPGSSPSVQCRSPRGKNIQGKTLVSQLELDPSG 180
QY 181 TWCTVQLONOKVEFKIDIV 200
DB 181 TWCTVQLONOKVEFKIDIV 200

RESULT 42

US-09-218-950-4
Sequence 4, Application US/09218950
Patent No. 6284240

GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/218,950
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991

ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-218-950-4

Query March 78.1% Score 1029; DB 3; Length 575;
Best Local Similarity 100.0%; Pred. No. 4.6e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVQLALLPAAQGNKVLGKKGDVETLTCTASOKKSIQFHKNSNQIK 60
DB 1 MNRGVPRHLLVQLALLPAAQGNKVLGKKGDVETLTCTASOKKSIQFHKNSNQIK 60
QY 61 ILGNQGSFLTKGPKSKNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPKSKNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDBTHLQGSQSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDBTHLQGSQSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTVLQONOKKVEFKIDIV 200
DB 181 TWCTVLQONOKKVEFKIDIV 200

RESULT 43
PCT-US92-01785-4
Sequence 4, Application PC/TUS9201785
GENERAL INFORMATION:
APPLICANT: The General Hospital Corporation
TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor
TITLE OF INVENTION: Chimeras
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/01785
FILING DATE: 19920306
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear

MOLECULE TYPE: protein
PCT-US92-01785-4

Query March 78.1% Score 1029; DB 5; Length 575;
Best Local Similarity 100.0%; Pred. No. 4.6e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVQLALLPAAQGNKVLGKKGDVETLTCTASOKKSIQFHKNSNQIK 60
DB 1 MNRGVPRHLLVQLALLPAAQGNKVLGKKGDVETLTCTASOKKSIQFHKNSNQIK 60
QY 61 ILGNQGSFLTKGPKSKNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPKSKNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDBTHLQGSQSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDBTHLQGSQSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTVLQONOKKVEFKIDIV 200
DB 181 TWCTVLQONOKKVEFKIDIV 200

RESULT 44
PCT-US95-00454-4
Sequence 4, Application PC/TUS9500454
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Targeted Cytolysis of HIV-Infected
TITLE OF INVENTION: Cells by Chimeric CD4 Receptor-
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00454
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-00454-4

Query Match 78.1%; Score 1029; DB 5; Length 575;
Best Local Similarity 100.0%; Pred. No. 4.6e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLILVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNSQIK 60
DB 1 MNRGVPFRHLILVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNSQIK 60

QY 61 ILGNQGSFLTKGSPKLNDRADSRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGSPKLNDRADSRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120

QY 121 LVFGLTANSPDTHLLQGSGLTLTLSPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPDTHLLQGSGLTLTLSPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180

QY 181 TWICTVLQNOKKEVFKIDIV 200
DB 181 TWICTVLQNOKKEVFKIDIV 200

RESULT 45
US-08-472-888A-6
Sequence 6, Application US/08472888A
Patent No. 6613746
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Walz, Gerd
TITLE OF INVENTION: AGP-ANTIBODY FUSION PROTEINS
TITLE OF INVENTION: AND RELATED MOLECULES AND METHODS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472.888A
CLASSIFICATION: 424
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/618,314
FILING DATE: 23-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/258001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-472-888A-6

Query Match 78.1%; Score 1029; DB 4; Length 630;
Best Local Similarity 100.0%; Pred. No. 5.2e-79;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLILVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNSQIK 60
DB 1 MNRGVPFRHLILVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNSQIK 60

DB 1 MNRGVPFRHLILVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNSQIK 60

QY 61 ILGNQGSFLTKGSPKLNDRADSRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGSPKLNDRADSRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120

QY 121 LVFGLTANSPDTHLLQGSGLTLTLSPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPDTHLLQGSGLTLTLSPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180

QY 181 TWICTVLQNOKKEVFKIDIV 200
DB 181 TWICTVLQNOKKEVFKIDIV 200

RESULT 46
5223394-9
Patent No. 5223394
APPLICANT: WALLNER, BARBARA
TITLE OF INVENTION: RECOMBINANT DNA MOLECULE COMPRISING
LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN 3 PHOSPHATIDYLINOSITOL
LINKAGE SIGNAL SEQUENCE
NUMBER OF SEQUENCES: 12
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/335,688
FILING DATE: 10-APR-1989
SEQ ID NO: 9
LENGTH: 295

Query Match 77.7%; Score 1023; DB 6; Length 295;
Best Local Similarity 99.5%; Pred. No. 6.1e-79;
Matches 199; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLILVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNSQIK 60
DB 1 MNRGVPFRHLILVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNSQIK 60

QY 61 ILGNQGSFLTKGSPKLNDRADSRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGSPKLNDRADSRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120

QY 121 LVFGLTANSPDTHLLQGSGLTLTLSPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPDTHLLQGSGLTLTLSPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180

QY 181 TWICTVLQNOKKEVFKIDIV 200
DB 181 TWICTVLQNOKKEVFKIDIV 200

RESULT 47
5223394-11
Patent No. 5223394
APPLICANT: WALLNER, BARBARA
TITLE OF INVENTION: RECOMBINANT DNA MOLECULE COMPRISING
LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN 3 PHOSPHATIDYLINOSITOL
LINKAGE SIGNAL SEQUENCE
NUMBER OF SEQUENCES: 12
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/335,688
FILING DATE: 10-APR-1989
SEQ ID NO: 11
LENGTH: 318

Query Match 77.7%; Score 1023; DB 6; Length 318;
Best Local Similarity 99.5%; Pred. No. 6.7e-79;
Matches 199; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLILVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNSQIK 60
DB 1 MNRGVPFRHLILVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNSQIK 60

QY 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLOQSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSVQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSVQLELDQSG 180
QY 181 TWTCVLOQKQVEFKIDIV 200
DB 181 TWTCVLOQKQVEFKIDIV 200

RESULT 48
US-09-039-555B-15
; Sequence 15, Application US/09039555B
; Patent No. 6033856
; GENERAL INFORMATION:
; APPLICANT: Koerner, Kathrin
; APPLICANT: Mueller, Rolf
; APPLICANT: Sadiacek, Hans-Harald
; TITLE OF INVENTION: PROMOTER OF THE CDC25B GENE, ITS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/039,555B
; FILING DATE: 16-MAR-1998
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 19710643.9
; FILING DATE: 14-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016779/0131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 458 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-039-555B-15

Query Match 77.7%; Score 1023; DB 3; Length 458;
Best Local Similarity 99.5%; Pred. No. 1.1e-78;
Matches 199; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVQLALPAATQGNKVVLGKGDVVELCTASQKSIQFHMKNNOIK 60
DB 1 MNRGVPRHLLVQLALPAATQGNKVVLGKGDVVELCTASQKSIQFHMKNNOIK 60
QY 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120

QY 121 LVFGLTANSDTHLLOQSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSVQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSVQLELDQSG 180
QY 181 TWTCVLOQKQVEFKIDIV 200
DB 181 TWTCVLOQKQVEFKIDIV 200

RESULT 49
US-09-517-605-3
; Sequence 3, Application US/09517605
; Patent No. 6391567
; GENERAL INFORMATION:
; APPLICANT: Litman, Dan R.
; APPLICANT: Kwon, Douglas S.
; APPLICANT: van Kooyk, Yvette
; APPLICANT: Geljtenbeck, Theo
; TITLE OF INVENTION: METHODS OF USING A FACILITATOR OF RETROVIRAL ENTRY INTO
; FILE REFERENCE: 1049-1-017
; CURRENT APPLICATION NUMBER: US/09/517,605
; CURRENT FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 3
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-517-605-3

Query Match 77.7%; Score 1023; DB 4; Length 458;
Best Local Similarity 99.5%; Pred. No. 1.1e-78;
Matches 199; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVQLALPAATQGNKVVLGKGDVVELCTASQKSIQFHMKNNOIK 60
DB 1 MNRGVPRHLLVQLALPAATQGNKVVLGKGDVVELCTASQKSIQFHMKNNOIK 60
QY 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLOQSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSVQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSVQLELDQSG 180
QY 181 TWTCVLOQKQVEFKIDIV 200
DB 181 TWTCVLOQKQVEFKIDIV 200

RESULT 50
5223394-7
; Patent No. 5223394
; APPLICANT: WALINER, BARBARA
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULE COMPRISING
; LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN 3 PHOSPHATIDYLINOSITOL
; LINKAGE SIGNAL SEQUENCE
; NUMBER OF SEQUENCES: 12
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/335,688
; FILING DATE: 10-APR-1989
; SEQ ID NO: 7
; LENGTH: 458
5223394-7

Query Match 77.1%; Score 1015; DB 6; Length 458;
Best Local Similarity 99.0%; Pred. No. 5.2e-78;
Matches 198; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVQLALPAATQGNKVVLGKGDVVELCTASQKSIQFHMKNNOIK 60
DB 1 MNRGVPRHLLVQLALPAATQGNKVVLGKGDVVELCTASQKSIQFHMKNNOIK 60

Db 1 MNRGVPFRHLLVLQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
QY 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIIRNLKIEDSDTYICEVEDQKEEVOL 120
Db 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIIRNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSPTHTLLQGSGLTLTLESPPGSSPVQCRSPRGKNIQGGKTLISVSOLELQDSG 180
Db 121 LVFGLTANSPTHTLLQGSGLTLTLESPPGSSPVQCRSPRGKNIQGGKTLISVSOLELQDSG 180
QY 181 TWICTVLQONOKKVEFKIDIV 200
Db 181 TWICTVLQONOKKVEFKIDIV 200

RESULT 51
US-08-466-368-2
; Sequence 2, Application US/08466368
; Patent No. 6093539
; GENERAL INFORMATION:
; APPLICANT: Maddon, Paul J.
; APPLICANT: Littman, Dan R.
; APPLICANT: Chess, Leonard
; APPLICANT: Axel, Richard
; APPLICANT: Weiss, Robin
; APPLICANT: McDougal, J. S.
; TITLE OF INVENTION: DNA ENCODING THE T CELL SURFACE PROTEIN
; TITLE OF INVENTION: T4 AND USE OF FRAGMENTS OF T4 IN THE TREATMENT OF AIDS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,368
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 24577-E1-B/JPM/AKC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 394 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-466-368-2

Query Match 76.8%; Score 1012; DB 3; Length 394;
Best Local Similarity 98.5%; Pred. No. 7.6e-78;
Matches 197; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLLVLQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
Db 1 MNRGVPFRHLLVLQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
QY 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIIRNLKIEDSDTYICEVEDQKEEVOL 120
Db 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIIRNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSPTHTLLQGSGLTLTLESPPGSSPVQCRSPRGKNIQGGKTLISVSOLELQDSG 180
Db 121 LVFGLTANSPTHTLLQGSGLTLTLESPPGSSPVQCRSPRGKNIQGGKTLISVSOLELQDSG 180
QY 181 TWICTVLQONOKKVEFKIDIV 200
Db 181 TWICTVLQONOKKVEFKIDIV 200

Db 121 LVFGLTANSPTHTLLQGSGLTLTLESPPGSSPVQCRSPRGKNIQGGKTLISVSOLELQDSG 180
QY 181 TWICTVLQONOKKVEFKIDIV 200
Db 181 TWICTVLQONOKKVEFKIDIV 200

RESULT 52
US-08-328-500-2
; Sequence 2, Application US/08328500
; Patent No. 6673896
; GENERAL INFORMATION:
; APPLICANT: Maddon, Paul J.
; APPLICANT: Sweet, Richard
; APPLICANT: Axel, Richard W.
; APPLICANT: Arthos, James
; TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,500
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/24577-CY
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 394 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-500-2

Query Match 76.8%; Score 1012; DB 4; Length 394;
Best Local Similarity 98.5%; Pred. No. 7.6e-78;
Matches 197; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLLVLQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
Db 1 MNRGVPFRHLLVLQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
QY 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIIRNLKIEDSDTYICEVEDQKEEVOL 120
Db 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIIRNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSPTHTLLQGSGLTLTLESPPGSSPVQCRSPRGKNIQGGKTLISVSOLELQDSG 180
Db 121 LVFGLTANSPTHTLLQGSGLTLTLESPPGSSPVQCRSPRGKNIQGGKTLISVSOLELQDSG 180
QY 181 TWICTVLQONOKKVEFKIDIV 200
Db 181 TWICTVLQONOKKVEFKIDIV 200

RESULT 53
5223418-2

```

; Patent No. 5223418
; APPLICANT: ARCURI, EDWARD J.; BRANNER, MARY E.; DONOVAN, MARY
; J.; GERBER, ROBERT G.; KELLER, JOHN A.
; TITLE OF INVENTION: METHOD OF IMPROVING THE YIELD OF
; HETEROLOGOUS PROTEINS PRODUCED BY STREPTOMYCES LIVIDANS
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/589,979
; FILING DATE: 28-SEP-1990
; SEQ ID NO: 2
; LENGTH: 394
5223418-2

Query Match          76.2%; Score 1003; DB 6; Length 394;
Best Local Similarity 98.0%; Pred. No. 4,4e-77;
Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLIVLQALLPATGKNVVLGKGDVLTCTASOKKSIOFHMKNNOIK 60
DB 1 MNRGVPFRHLIVLQALLPATGKVVLGKGDVLTCTASOKKSIOFHMKNNOIK 60
QY 61 ILGNQGSFLTKGPKLNDRADSRRLWDQGNFLLIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPKLNDRADSRRLWDQGNFLLIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTNSDTHLLOQSITLTLESPPGSSPSVQCRPRGKNTGGKTLVSQLELDSG 180
DB 121 LVFGLTNSDTHLLOQSITLTLESPPGSSPSVQCRPRGKNTGGKTLVSQLELDSG 180
QY 181 TWTCTVLONOKKVEFKIDIV 200
DB 181 TWTCTVLONOKKVEFKIDIV 200

```

```

RESULT 54
US-08-236-311-4
; Sequence 4, Application US/08236311
; Patent No. 5565335
; GENERAL INFORMATION:
; APPLICANT: Capon, Daniel J.
; APPLICANT: Gregory, Timothy J.
; TITLE OF INVENTION: Adhesion Variants
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/236,311
; FILING DATE: 02-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/936190
; FILING DATE: 26-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/842777
; FILING DATE: 18-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/250785
; FILING DATE: 28-SEP-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/104329
; FILING DATE: 02-OCT-1987
; ATTORNEY/AGENT INFORMATION:
; NAME: Haasek, Janet E.

```

```

; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 444P1C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 434 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-236-311-4

Query Match          69.6%; Score 916; DB 1; Length 434;
Best Local Similarity 100.0%; Pred. No. 1.2e-69;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 QGNKVVLGKGDVLTCTASOKKSIOFHMKNNOIKILGNQGSFLTKGPKLNDRADSR 83
DB 56 QGNKVVLGKGDVLTCTASOKKSIOFHMKNNOIKILGNQGSFLTKGPKLNDRADSR 115
QY 84 RSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVOLLVFGLTNSDTHLLOQSITLT 143
DB 116 RSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVOLLVFGLTNSDTHLLOQSITLT 175
QY 144 ESPPGSSPSVQCRPRGKNTGGKTLVSQLELDSGTTCTVLONOKKVEFKIDIV 200
DB 176 ESPPGSSPSVQCRPRGKNTGGKTLVSQLELDSGTTCTVLONOKKVEFKIDIV 232

```

```

RESULT 55
US-08-457-918-4
; Sequence 4, Application US/08457918
; Patent No. 6117655
; GENERAL INFORMATION:
; APPLICANT: Capon, Daniel J.
; APPLICANT: Gregory, Timothy J.
; TITLE OF INVENTION: Adhesion Variants
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,918
; FILING DATE: 1-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/236311
; FILING DATE: 02-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/936190
; FILING DATE: 26-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/842777
; FILING DATE: 18-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/250785
; FILING DATE: 28-SEP-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/104329
; FILING DATE: 02-OCT-1987
; ATTORNEY/AGENT INFORMATION:
; NAME: Kudinec, Jeffrey S.
; REGISTRATION NUMBER: 36,575

```


REFERENCE/DOCKET NUMBER: P0444P1C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-8228
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 434 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-457-918-4

Query Match 69.6%; Score 916; DB 3; Length 434;
Best Local Similarity 100.0%; Pred. No. 1,2e-69;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 QGNKVLGKGDVLTCTASQKSIQFHMKNNOIKILNGSGFLTKGPKLNDRADSR 83
DB 56 QGNKVLGKGDVLTCTASQKSIQFHMKNNOIKILNGSGFLTKGPKLNDRADSR 115
QY 84 RSLMDQGNFPLIIKNLIKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLLOQGSLLTLES 143
DB 116 RSLMDQGNFPLIIKNLIKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLLOQGSLLTLES 175
QY 144 ESPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSGTWCTVLOKQKVEFKIDIV 200
DB 176 ESPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSGTWCTVLOKQKVEFKIDIV 232

RESULT 56
US-08-630-172-17
Sequence 17, Application US/08630172
Patent No. 6060054
GENERAL INFORMATION:
APPLICANT: Staerz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630.172
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-0223
TELEFAX: (303) 863-9700
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-630-172-17

Query Match 69.4%; Score 914; DB 3; Length 410;
Best Local Similarity 79.5%; Pred. No. 1.6e-69;
Matches 186; Conservative 9; Mismatches 27; Indels 12; Gaps 4;

QY 26 NKVVLGKGDVLTCTASQKSIQFHMKNNOIKILNGSGFLTKGPKLNDRADSR 85
DB 1 NKVVLGKGDVLTCTASQKSIQFHMKNNOIKILNGSGFLTKGPKLNDRADSR 60
QY 86 LMDQGNFPLIIKNLIKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLLOQGSLLTLES 145
DB 61 LMDQGNFPLIIKNLIKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLLOQGSLLTLES 120
QY 146 PGSSSPVQCRSPRGKNIQGGKTLVSQLELDQSGTWCTVLOKQKVEFKIDIV----P 201
DB 121 PGSSSPVQCRSPRGKNIQGGKTLVSQLELDQSGTWCTVLOKQKVEFKIDIVLAEP 180
QY 202 RASAL-PAPPTGSLPDPQTASALPDPASALPALAVLSFLGLGACVLA 254
DB 181 RGPITKCPCKCAPN-----LLGSPSVFIIPPRIKIV-LMISLPIYTCV 227

RESULT 57
US-09-375-419-17
Sequence 17, Application US/09375419
Patent No. 6264950
GENERAL INFORMATION:
APPLICANT: Staerz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/375.419
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630.172
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-0223
TELEFAX: (303) 863-9700
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-375-419-17

Query Match 69.4%; Score 914; DB 3; Length 410;
Best Local Similarity 79.5%; Pred. No. 1.6e-69;
Matches 186; Conservative 9; Mismatches 27; Indels 12; Gaps 4;

QY 26 NKVVLGKGDVLTCTASQKSIQFHMKNNOIKILNGSGFLTKGPKLNDRADSR 85
DB 1 NKVVLGKGDVLTCTASQKSIQFHMKNNOIKILNGSGFLTKGPKLNDRADSR 60
QY 86 LMDQGNFPLIIKNLIKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLLOQGSLLTLES 145
DB 61 LMDQGNFPLIIKNLIKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLLOQGSLLTLES 120
QY 146 PGSSSPVQCRSPRGKNIQGGKTLVSQLELDQSGTWCTVLOKQKVEFKIDIV----P 201

Db 121 PGGSSPVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLOKQKVEFKIDIVLAEP 180
Qy 202 RASAL-PAPPTGSAIPDPTASALPDPPAASALPALAVISFLGLGAGVAVL 254
181 RGPTRKCPCKPCAPFN-----LLGGPSVFIFPPKIDV-LMISLSPIYTCV 227

RESULT 58
US-09-100-409A-1
; Sequence 1, Application US/09100409A
; Patent No. 6090388
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
; TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND
; TITLE OF INVENTION: IMMUNE DISORDERS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100.409A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 1151-4154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-100-409A-1

Query Match 68.3%; Score 899; DB 3; Length 433;
Best Local Similarity 100.0%; Pred. No. 3.2e-68;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 KVVIGKKGDTVELCTASQKSIQFHMKNNOIKILGNQSGFLTKGPSKLNDRADSRSL 86
Db 2 KVVIGKKGDTVELCTASQKSIQFHMKNNOIKILGNQSGFLTKGPSKLNDRADSRSL 61
Qy 87 WDOGNFPLIIKNIKIEDSDTYICEVEDQKEVQLVFGILTANSDTHLLOQGSLLTLLESP 146
Db 62 WDOGNFPLIIKNIKIEDSDTYICEVEDQKEVQLVFGILTANSDTHLLOQGSLLTLLESP 121
Qy 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLOKQKVEFKIDIV 200
Db 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLOKQKVEFKIDIV 175

RESULT 59
US-08-021-601-12
; Sequence 12, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppa, Stephen H.

APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021.601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-12

Query Match 68.3%; Score 899; DB 1; Length 903;
Best Local Similarity 100.0%; Pred. No. 8.5e-68;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 KVVIGKKGDTVELCTASQKSIQFHMKNNOIKILGNQSGFLTKGPSKLNDRADSRSL 86
Db 727 KVVIGKKGDTVELCTASQKSIQFHMKNNOIKILGNQSGFLTKGPSKLNDRADSRSL 786
Qy 87 WDOGNFPLIIKNIKIEDSDTYICEVEDQKEVQLVFGILTANSDTHLLOQGSLLTLLESP 146
Db 787 WDOGNFPLIIKNIKIEDSDTYICEVEDQKEVQLVFGILTANSDTHLLOQGSLLTLLESP 846
Qy 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLOKQKVEFKIDIV 200
Db 847 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLOKQKVEFKIDIV 900

RESULT 60
US-08-082-849B-12
; Sequence 12, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppa, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Atora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California

```

; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082,849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 903 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-082-849B-12

Query Match      68.3%; Score 899; DB 1; Length 903;
Best Local Similarity 100.0%; Pred. No. 8,5e-68;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 KVLGKGGDTVELTCTASQKKSIFHWKNSNQIKILGQGSFLTKGPSKINDRADSRSL 86
      |||||
DB 727 KVLGKGGDTVELTCTASQKKSIFHWKNSNQIKILGQGSFLTKGPSKINDRADSRSL 786
      |||||

QY 87 WDCGNFPLIKNKIEEDPTICEVEDQKEVQLLVFGLTANSDTHLLOQSLLTLESP 146
      |||||
DB 787 WDCGNFPLIKNKIEEDPTICEVEDQKEVQLLVFGLTANSDTHLLOQSLLTLESP 846
      |||||

QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTCYLQNKVKEFKIDIV 200
      |||||
DB 847 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTCYLQNKVKEFKIDIV 900
      |||||

RESULT 61
PCT-US94-01624-12
; Sequence 12, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nicholas, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND KHOURI and CREW
; STREET: Steuart Street Tower, 20th floor, One Market
; STREET: Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01624
```

```

; FILING DATE: June 25, 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-115
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 903 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-01624-12

Query Match      68.3%; Score 899; DB 5; Length 903;
Best Local Similarity 100.0%; Pred. No. 8,5e-68;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 KVLGKGGDTVELTCTASQKKSIFHWKNSNQIKILGQGSFLTKGPSKINDRADSRSL 86
      |||||
DB 727 KVLGKGGDTVELTCTASQKKSIFHWKNSNQIKILGQGSFLTKGPSKINDRADSRSL 786
      |||||

QY 87 WDCGNFPLIKNKIEEDPTICEVEDQKEVQLLVFGLTANSDTHLLOQSLLTLESP 146
      |||||
DB 787 WDCGNFPLIKNKIEEDPTICEVEDQKEVQLLVFGLTANSDTHLLOQSLLTLESP 846
      |||||

QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTCYLQNKVKEFKIDIV 200
      |||||
DB 847 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTCYLQNKVKEFKIDIV 900
      |||||
```

```

RESULT 62
US-08-867-149-1
; Sequence 1, Application US/08867149
; Patent No. 5912176
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: Antibody Against a Host Cell
; TITLE OF INVENTION: Antigen Complex for Pre- and Post-Exposure
; TITLE OF INVENTION: Protection from Infection by HIV Primary Isolates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,149
; FILING DATE: 28-Feb-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/657,149
; FILING DATE: 03-June-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C.H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4145
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)415-8745
; TELEFAX: (212)751-6849
; INFORMATION FOR SEQ ID NO: 1:
```

SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-867-149-1

Query Match 68.2%; Score 898; DB 2; Length 433;
Best Local Similarity 99.4%; Pred. No. 3,9e-68;
Matches 174; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 26 NKVLGKKGDTVELTCTASQKSIQFHMKNNOIKILGNQGSFLTKGPSKLNDRADSRSS 85
DB 1 NKVLGKKGDTVELTCTASQKSIQFHMKNNOIKILGNQGSFLTKGPSKLNDRADSRSS 60
QY 86 LMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLVFGILTANSDTHLQGSFLTLES 145
DB 61 LMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLVFGILTANSDTHLQGSFLTLES 120
QY 146 PGSSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLQNKVKVERKIDIV 200
DB 121 PGSSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLQNKVKVERKIDIV 175

RESULT 63
US-08-808-374-1
Sequence 1, Application US/08808374
Patent No. 5961976

GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: Antibody Against a Host Cell
TITLE OF INVENTION: Antigen Complex for Pre- and Post-Exposure
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Maria C.H. Lin
STREET: 345 Park Avenue
CITY: New York
STATE: NY

COUNTRY: USA

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version

SOFTWARE: #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808,374

FILING DATE: 28-Feb-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/657,149

FILING DATE: 03-June-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4145

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)415-8745

TELEFAX: (212)751-6849

INFORMATION FOR SEQ ID NO. 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-808-374-1

Query Match 68.2%; Score 898; DB 2; Length 433;
Best Local Similarity 99.4%; Pred. No. 3,9e-68;
Matches 174; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 26 NKVLGKKGDTVELTCTASQKSIQFHMKNNOIKILGNQGSFLTKGPSKLNDRADSRSS 85
DB 1 NKVLGKKGDTVELTCTASQKSIQFHMKNNOIKILGNQGSFLTKGPSKLNDRADSRSS 60
QY 86 LMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLVFGILTANSDTHLQGSFLTLES 145
DB 61 LMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLVFGILTANSDTHLQGSFLTLES 120
QY 146 PGSSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLQNKVKVERKIDIV 200
DB 121 PGSSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLQNKVKVERKIDIV 175

RESULT 64
5171838-13
Patent No. 5171838
APPLICANT: CHIBA, YUKIOBU
TITLE OF INVENTION: LEU3 BINDING PEPTIDES
NUMBER OF SEQUENCES: 24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/526,921
FILING DATE: 22-MAY-1990
SEQ ID NO:13:
LENGTH: 433
5171838-13

Query Match 54.5%; Score 718; DB 6; Length 433;
Best Local Similarity 83.4%; Pred. No. 6,9e-53;
Matches 151; Conservative 6; Mismatches 10; Indels 14; Gaps 4;

QY 27 KVLGKKGDTVELTCTASQKSIQFHMKNNOIKILGNQGSFLTKGPSKLNDRADSRSS- 85
DB 2 KVLGKKGDTVELTCTASQKSIQFHMKNNOIKILGNQGSFLTKGPSKLNDRADSRSS 61
QY 86 ----LMDQGNFPLIKNLKIED-SDTYICEVEDQKEEVQLVFGILTANSDTHLQGSFL 139
DB 62 NQIKILGNQGSF-LTGPSKLNDRADS-----RSEEVQLVFGILTANSDTHLQGSFL 114
QY 140 TLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLQNKVKVERKIDIV 199
DB 115 TLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLQNKVKVERKIDIV 174
QY 200 V 200
DB 175 V 175

RESULT 65
US-08-630-172-1
Sequence 1, Application US/08630172
Patent No. 6060054

GENERAL INFORMATION:
APPLICANT: Staerz, Uwe

TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T

TITLE OF INVENTION: LYMPHOCYTE VETO

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh

STREET: 1700 Lincoln Street, 35th Floor

CITY: Denver

STATE: Colorado

COUNTRY: U.S.

ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,172

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 138 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-630-172-1

Query Match 54.4%; Score 716; DB 3; Length 138;
Best Local Similarity 100.0%; Pred. No. 2,2e-53;
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 NKVLGKKGDVVELTCTASQKSIQFHWKNSNQIKILGNQGSFLTKGPKLNDRAISR 85
DB 1 NKVLGKKGDVVELTCTASQKSIQFHWKNSNQIKILGNQGSFLTKGPKLNDRAISR 60
QY 86 LMDGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANS DTHLHOGS LTLTLES 145
DB 61 LMDGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANS DTHLHOGS LTLTLES 120
QY 146 PGSSPSVQCRSPRGKNI 163
DB 121 PGSSPSVQCRSPRGKNI 138

RESULT 66
US-09-375-419-1
Sequence 1, Application US/09375419
Patent No. 6264950
GENERAL INFORMATION:
APPLICANT: Staerz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
TITLE OF INVENTION: LYMPHOCYTE VETO
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/375,419
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,172
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 138 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-09-375-419-1

Query Match 54.4%; Score 716; DB 3; Length 138;
Best Local Similarity 100.0%; Pred. No. 2,2e-53;
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 NKVLGKKGDVVELTCTASQKSIQFHWKNSNQIKILGNQGSFLTKGPKLNDRAISR 85
DB 1 NKVLGKKGDVVELTCTASQKSIQFHWKNSNQIKILGNQGSFLTKGPKLNDRAISR 60
QY 86 LMDGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANS DTHLHOGS LTLTLES 145
DB 61 LMDGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANS DTHLHOGS LTLTLES 120
QY 146 PGSSPSVQCRSPRGKNI 163
DB 121 PGSSPSVQCRSPRGKNI 138

RESULT 67
US-08-558-269-10
Sequence 10, Application US/08558269
Patent No. 5961973
GENERAL INFORMATION:
APPLICANT: Crea, Roberto
TITLE OF INVENTION: PATHOGEN-TARGETED BIOCATALYSTS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/558,269
FILING DATE: 13-NOV-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/847,800
FILING DATE: 06-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: CRI-001CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-558-269-10

Query Match 45.2%; Score 595.5; DB 2; Length 376;
Best Local Similarity 92.2%; Pred. No. 1,4e-42;
Matches 118; Conservative 1; Mismatches 6; Indels 3; Gaps 1;

QY 27 KVLGKKGDVVELTCTASQKSIQFHWKNSNQIKILGNQGSFLTKGPKLNDRAISR 86
DB 3 KVLGKKGDVVELTCTASQKSIQFHWKNSNQIKILGNQGSFLTKGPKLNDRAISR 62
QY 87 WDOGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANS DTHLHOGS LTLTLES 146
DB 63 WDOGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANS DTHLHOGS LTLTLES 122

QY 147 P---GSSP 151
DB 123 DAIGMSP 130

RESULT 68

US-09-410-882-10
; Sequence 10, Application US/09410882
; Patent No. 6287561
; GENERAL INFORMATION:
; APPLICANT: Crea, Roberto
; TITLE OF INVENTION: PATHOGEN-TARGETED BIOCATALYSTS
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/410,882
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/558,269
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: US 07/847,800
; FILING DATE: 06-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: CRI-001CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-410-882-10

Query Match 45.2%; Score 595.5; DB 3; Length 376;
Best Local Similarity 92.2%; Pred. No. 1.4e-42;
Matches 118; Conservative 1; Mismatches 6; Indels 3; Gaps 1;

QY 27 KVLGKKGDVTELTCTASQKKSIOFHWKNSNOIKILGNQSFITKGPSKLNDRADSRSL 86
DB 3 KVLGKKGDVTELTCTASQKKSIOFHWKNSNOIKILGNQSFITKGPSKLNDRADSRSL 62
QY 87 WDOGNFPLIIKNIKIEDSDTYICEVEDQKEVQLVFGLTANSPTHLHOGQSLTLTLESP 146
DB 63 WDOGNFPLIIKNIKIEDSDTYICEVEDQKEVQLVFGLTANSPTHLHOGQSLTLTLEGS 122
QY 147 P---GSSP 151
DB 123 DAIGMSP 130

RESULT 69

US-08-558-269-6
; Sequence 6, Application US/08558269
; Patent No. 5961973
; GENERAL INFORMATION:
; APPLICANT: Crea, Roberto
; TITLE OF INVENTION: PATHOGEN-TARGETED BIOCATALYSTS

NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/558,269
FILING DATE: 13-NOV-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/847,800
FILING DATE: 06-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: CRI-001CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 383 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-558-269-6

Query Match 44.8%; Score 589.5; DB 2; Length 383;
Best Local Similarity 72.9%; Pred. No. 4.5e-42;
Matches 121; Conservative 9; Mismatches 23; Indels 13; Gaps 2;

QY 27 KVLGKKGDVTELTCTASQKKSIOFHWKNSNOIKILGNQSFITKGPSKLNDRADSRSL 86
DB 3 KVLGKKGDVTELTCTASQKKSIOFHWKNSNOIKILGNQSFITKGPSKLNDRADSRSL 62
QY 87 WDOGNFPLIIKNIKIEDSDTYICEVEDQKEVQLVFGLTANSPTHLHOGQSLTLTLESP 146
DB 63 WDOGNFPLIIKNIKIEDSDTYICEVEDQKEVQLVFGLTANSPTHLHOGQSLTLTLEGS 117
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLEIADSGTWTCTVLQNOKR 192
DB 118 --STCGLRQYSQOPRIKQGLFADIA-----SHPQAAIIFAGRR 155

RESULT 70

US-09-410-882-6
; Sequence 6, Application US/09410882
; Patent No. 6287561
; GENERAL INFORMATION:
; APPLICANT: Crea, Roberto
; TITLE OF INVENTION: PATHOGEN-TARGETED BIOCATALYSTS
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/410,882
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/558,269
FILING DATE: 13-NOV-1995
APPLICATION NUMBER: US 07/847,800
FILING DATE: 06-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: CRI-001CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 383 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-410-882-6

Query Match 44.8%; Score 589.5; DB 3; Length 383;
Best Local Similarity 72.9%; Pred. No. 4,5e-42;
Matches 121; Conservative 9; Mismatches 23; Indels 13; Gaps 2;

QY 27 KVVIGKKGDVVELTCTASQKKSIOFHMKNSQIKILNQGSFLTKGSKLNDRADSRSL 86
DB 3 KVVIGKKGDVVELTCTASQKKSIOFHMKNSQIKILNQGSFLTKGSKLNDRADSRSL 62

QY 87 WDQGNFPLIINKLKIEDSDTYICEVEDQKEVQLVGLTANSTHLLQGSLLTLTLESP 146
DB 63 WDQGNFPLIINKLKIEDSDTYICEVEDQKEVQLVGLTANSTHLLQGSLLTLTLESP 117

QY 147 PGSSPSVQCRSPRGKNIQGGKTLISVSOLELDQSGTWTCTVIONOKK 192
DB 118 --STCGLRQYSQPOFRKIGGLFADIA-----SHPMQALPAKRR 155

RESULT 71
US-08-466-368-5
Sequence 5, Application US/08466368
Patent No. 6093539
GENERAL INFORMATION:
APPLICANT: Maddon, Paul J.
APPLICANT: Littman, Dan R.
APPLICANT: Chess, Leonard
APPLICANT: Axel, Richard
APPLICANT: Weles, Robin
APPLICANT: McDougal, J. S.
TITLE OF INVENTION: DNA ENCODING THE T CELL SURFACE PROTEIN
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,368
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 24577-BI-B/JPW/AKC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 94 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: YES
FEATURE:
NAME/KEY: Active-site
LOCATION: 1..94
US-08-466-368-5

Query Match 37.5%; Score 494; DB 3; Length 94;
Best Local Similarity 100.0%; Pred. No. 8,5e-35;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 QGNKVVIGKKGDVVELTCTASQKKSIOFHMKNSQIKILNQGSFLTKGSKLNDRADSR 83
DB 1 QGNKVVIGKKGDVVELTCTASQKKSIOFHMKNSQIKILNQGSFLTKGSKLNDRADSR 60

QY 84 RSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEE 117
DB 61 RSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEE 94

RESULT 72
US-08-470-998-2
Sequence 2, Application US/08470998
Patent No. 6570000
GENERAL INFORMATION:
APPLICANT: Maddon, Paul J.
APPLICANT: Littman, Dan R.
APPLICANT: Chess, Leonard
APPLICANT: Axel, Richard
APPLICANT: Weles, Robin
APPLICANT: McDougal, J. S.
TITLE OF INVENTION: DNA ENCODING THE T CELL SURFACE PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,998
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 24577-BI-B/JPW/AKC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 94 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: YES
FEATURE:
NAME/KEY: Active-site
LOCATION: 1..94
US-08-470-998-2

Query Match 37.5%; Score 494; DB 4; Length 94;
Best Local Similarity 100.0%; Pred. No. 8.5e-35;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

24 QGNKVVVGKGGDTVELTCTASOKSIOFHWKNSQIKILNQSFLTKGSKLNDRAISR 83
1 QGNKVVVGKGGDTVELTCTASOKSIOFHWKNSQIKILNQSFLTKGSKLNDRAISR 60

84 RSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEE 117
61 RSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEE 94

RESULT 73

US-08-328-500-10
Sequence 10; Application US/08328500
Patent No. 6673896
GENERAL INFORMATION:
APPLICANT: Maddon, Paul J.
APPLICANT: Axel, Richard
APPLICANT: Sweet, Richard W.
APPLICANT: Arthos, James
TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,500
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/24577-CY
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 94 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-08-328-500-10

Query Match 37.5%; Score 494; DB 4; Length 94;
Best Local Similarity 100.0%; Pred. No. 8.5e-35;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

24 QGNKVVVGKGGDTVELTCTASOKSIOFHWKNSQIKILNQSFLTKGSKLNDRAISR 83
1 QGNKVVVGKGGDTVELTCTASOKSIOFHWKNSQIKILNQSFLTKGSKLNDRAISR 60

84 RSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEE 117

61 RSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEE 94

RESULT 74
US-08-416-478A-8
Sequence 8; Application US/08416478A
Patent No. 5773578

GENERAL INFORMATION:
APPLICANT: Hercend, Thierry
APPLICANT: Tielebel, Frederic
TITLE OF INVENTION: New Proteins Produced By Human
Lymphocytes, DNA Sequences Encoding These Proteins And
Their Pharmaceutical And Biological Uses
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 400
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/416,478A
FILING DATE: 04-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,644
FILING DATE: 08-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: HERCEND=1A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 457 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-416-478A-8

Query Match 37.2%; Score 490.5; DB 1; Length 457;
Best Local Similarity 48.8%; Pred. No. 1.4e-33;
Matches 102; Conservative 35; Mismatches 67; Indels 5; Gaps 3;

1 MNRGVFFRLH--LLVQLALPLPATQGNKVVVGKGGDTVELTCTASOKSIOFHWKNSQ 58
1 MCRGFSFRHLPLLLQLSKLVLVYQKTVVLKREGSALPEESTSRSSASAMKSSDQ 60

59 IKILNQSFLTKGPKLNDRAISRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEY 118
61 KTLGYKNKLLIKGSLVSRFDSRKNAMERGSFPLINKLRMEDSGQTYCELENKEEY 120

119 QLLVFGILTANSDFHLIQSGSLTITLES-PPGSSPSVQCRSPRGKNIQGGKTLVSQLEIQ 177
121 ELWVFRVTNPGTRRLQGGSLTILIDSNSPKVSPPIECRKGSSNIVKDSKAFSTHSLRIQ 180

178 DSGTWCTVLONQKVEF--KIDIYPRAS 204
181 DSGIWNCTVTLNOKKHSFMDKLSVLGFS 209

RESULT 75
US-08-474-988B-8


```
; Sequence 8, Application US/08474988B
; Patent No. 5874250
; GENERAL INFORMATION:
; APPLICANT: Hercend, Thierry
; APPLICANT: Triebel, Frederic
; TITLE OF INVENTION: NEW PROTEINS PRODUCED BY HUMAN
; TITLE OF INVENTION: LYMPHOCYTES, DNA SEQUENCES ENCODING THESE PROTEINS AND
; TITLE OF INVENTION: THEIR PHARMACEUTICAL AND BIOLOGICAL USES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 Seventh Street, N.W., Suite 400
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,988B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; APPLICATION NUMBER: US 08/416,478
; FILING DATE: 04-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/854,644
; FILING DATE: 08-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: YUN, Allen C.
; REGISTRATION NUMBER: 37,971
; REFERENCE/DOCKET NUMBER: HERCEND=1B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 457 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-474-988B-8

Query Match      37.2%; Score 490.5; DB 2; Length 457;
Best Local Similarity 48.8%; Pred. No. 1.4e-33;
Matches 102; Conservative 35; Mismatches 67; Indels 5; Gaps 3;

QY 1 MNRGVPFRRL--LLVLQALLPAATQGNKVLGKKGDVETLTCTASQKSIQFHKNSQ 58
Db 1 MCRGFSFRHLPLLLQLSKLVLVTOGKTVLKGEGSALPCSTSRASFAWKSSDQ 60
QY 59 IKILNGSFLTKGPSKLNDRADRSRLMDQGNFPLIKVLTIEDSDTYICEVDQKEEV 118
Db 61 KTLGYKNKLLIKGSLSLYSRFSRKNAWERSFPLIINKLRMEDSQTVCCELENKKEEV 120
QY 119 QLLVFGILTANSDTHLLQGOSLTLTLES-PPGSSPSVOCSPRKNIGOGKTLVSOLEQ 177
Db 121 ELWVFRTFNPGRRLLLQGOSLTLTLDSPKVSDBPIECKKSSNIVDSKAFTHSLRIQ 180
QY 178 DSGTWCTVLQONQKVEF--KIDIVPRAS 204
Db 181 DSGIWNCTVTLLNKGKHSFDMKLSVLGFAS 209

RESULT 76
US-08-394-442B-8
; Sequence 8, Application US/08394442B
; Patent No. 5976877
```

```
; GENERAL INFORMATION:
; APPLICANT: Hercend, Thierry
; APPLICANT: Triebel, Frederic
; TITLE OF INVENTION: New Proteins Produced By Human
; TITLE OF INVENTION: Lymphocytes, DNA Sequences Encoding These Proteins And
; TITLE OF INVENTION: Their Pharmaceutical And Biological Uses
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 Seventh Street, N.W., Suite 400
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/394,442B
; FILING DATE: 24-FEB-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: HERCEND=2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 457 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-394-442B-8

Query Match      37.2%; Score 490.5; DB 2; Length 457;
Best Local Similarity 48.8%; Pred. No. 1.4e-33;
Matches 102; Conservative 35; Mismatches 67; Indels 5; Gaps 3;

QY 1 MNRGVPFRRL--LLVLQALLPAATQGNKVLGKKGDVETLTCTASQKSIQFHKNSQ 58
Db 1 MCRGFSFRHLPLLLQLSKLVLVTOGKTVLKGEGSALPCSTSRASFAWKSSDQ 60
QY 59 IKILNGSFLTKGPSKLNDRADRSRLMDQGNFPLIKVLTIEDSDTYICEVDQKEEV 118
Db 61 KTLGYKNKLLIKGSLSLYSRFSRKNAWERSFPLIINKLRMEDSQTVCCELENKKEEV 120
QY 119 QLLVFGILTANSDTHLLQGOSLTLTLES-PPGSSPSVOCSPRKNIGOGKTLVSOLEQ 177
Db 121 ELWVFRTFNPGRRLLLQGOSLTLTLDSPKVSDBPIECKKSSNIVDSKAFTHSLRIQ 180
QY 178 DSGTWCTVLQONQKVEF--KIDIVPRAS 204
Db 181 DSGIWNCTVTLLNKGKHSFDMKLSVLGFAS 209

RESULT 77
US-08-332-562A-84
; Sequence 84, Application US/08332562A
; Patent No. 5985599
; GENERAL INFORMATION:
; APPLICANT: MCKENZIE, Ian F. C.
; APPLICANT: HOGARTH, Mark P. C.
; APPLICANT: HIBBS, Margaret L.
; APPLICANT: SCOTT, Bernadette M.
; APPLICANT: BONADONNA, Lisa
; APPLICANT: HULETT, Mark D.
; TITLE OF INVENTION: FC RECEPTOR FOR IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 136
```

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Foley & Lardner
;; STREET: 3000 K Street, N.W., Suite 500
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20007-5109
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/332,562A
;; FILING DATE: 31-OCT-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/896,457
;; FILING DATE: 27-MAY-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BENT, Stephen A.
;; REGISTRATION NUMBER: 29,768
;; REFERENCE/DOCKET NUMBER: 54270/119/GRHA
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202)672-5300
;; TELEFAX: (202)672-5399
;; TELEX: 904136
;; INFORMATION FOR SEQ ID NO: 84:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 80 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-332-562A-84

Query Match 30.2%; Score 398; DB 2; Length 80;
Best Local Similarity 98.7%; Pred. No. 9.3e-27;
Matches 77; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 123 FGLTANSDTHLQGGSLITLLESPGSSPSVOCSPKGNIOGGKITLSVSELELDGSGTW 132
DB 1 FGLTANSDTHLQGGSLITLLESPGSSPSVOCSPKGNIOGGKITLSVSELELDGSGTW 60

QY 183 TCTVLQNKQKVEPKIDIV 200
DB 61 TCTVLQNKQKVEPKIDIV 78

RESULT 78
US-09-050-861B-8
; Sequence 8, Application US/09050861B
; Patent No. 655314
; GENERAL INFORMATION:
; APPLICANT: Payan, Donald
; TITLE OF INVENTION: TOSO AS A TARGET FOR DRUG SCREENING
; FILE REFERENCE: RIGL-002CON
; CURRENT APPLICATION NUMBER: US/09/050,861B
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: US/09/651,150B
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: US 09/050,861
; PRIOR FILING DATE: 1998-03-30
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 8
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-050-861B-8

Query Match 29.8%; Score 392; DB 4; Length 74;
Best Local Similarity 100.0%; Pred. No. 2.7e-26;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 VELTCTASOKKSIQFHWKNSNOIKLNGSFLTKGPSKLANDRADSRSLMDQGNPFLII 96
DB 1 VELTCTASOKKSIQFHWKNSNOIKLNGSFLTKGPSKLANDRADSRSLMDQGNPFLII 60

QY 97 KNLKIEDSDTYICE 110
DB 61 KNLKIEDSDTYICE 74

RESULT 79
US-09-100-409A-20
; Sequence 20, Application US/09100409A
; Patent No. 6030388
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
; PREVENTION AND TREATMENT OF HIV INFECTION AND
; TITLE OF INVENTION: IMMUNE DISORDERS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & PINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,409A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 1151-4154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: amino acid
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-100-409A-20

Query Match 23.8%; Score 314; DB 3; Length 61;
Best Local Similarity 100.0%; Pred. No. 8.5e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 SKINDRADSRSLMDQGNPFLIIKNLKIEDSDTYICEVEDQKEVOLLVFGLTANSQTHL 133
DB 1 SKINDRADSRSLMDQGNPFLIIKNLKIEDSDTYICEVEDQKEVOLLVFGLTANSQTHL 60

QY 134 L 134
DB 61 L 61

RESULT 80
US-08-284-391B-35
; Sequence 35, Application US/08284391B
; Patent No. 5651628
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Banapour, Babak
; APPLICANT: Romeo, Charles

```

; APPLICANT: Kolanus, Waldemar
; TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INJECTED
; TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,391B
; FILING DATE: 02-AUG-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/195,395
; FILING DATE: 14-FEB-1994
; APPLICATION NUMBER: 07/847,566
; FILING DATE: 06-MAR-1992
; APPLICATION NUMBER: 07/665,961
; FILING DATE: 07-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elbing, Karen L
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/247001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-284-391B-35

Query Match      21.9%; Score 288; DB 2; Length 58;
Best Local Similarity 100.0%; Pred. No. 1.3e-17;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      201 PRASALPAPPTGSLPDPQTASALPDPASALPALAVISFLIGLGVACVLAARR 258
DB      1 PRASALPAPPTGSLPDPQTASALPDPASALPALAVISFLIGLGVACVLAARR 58

RESULT 81
US-09-218-950-35
; Sequence 35, Application US/09218950
; Patent No. 6284240
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Banapour, Babak
; APPLICANT: Romeo, Charles
; APPLICANT: Kolanus, Waldemar
; TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
; TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/218,950
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,391
; FILING DATE: 02-AUG-1994
; APPLICATION NUMBER: 08/195,395
; FILING DATE: 14-FEB-1994
; APPLICATION NUMBER: 07/847,566
; FILING DATE: 06-MAR-1992
; APPLICATION NUMBER: 07/665,961
; FILING DATE: 07-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elbing, Karen L
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/247001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-218-950-35

Query Match      21.9%; Score 288; DB 3; Length 58;
Best Local Similarity 100.0%; Pred. No. 1.3e-17;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      201 PRASALPAPPTGSLPDPQTASALPDPASALPALAVISFLIGLGVACVLAARR 258
DB      1 PRASALPAPPTGSLPDPQTASALPDPASALPALAVISFLIGLGVACVLAARR 58

RESULT 82
US-09-100-409A-10
; Sequence 10, Application US/09100409A
; Patent No. 6090388
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang YI
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
; TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND
; TITLE OF INVENTION: IMMUNE DISORDERS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,409A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
```

REFERENCE/DOCKET NUMBER: 1151-4154
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 46 amino acids

TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: peptide
US-09-100-409A-10

Query Match 17.3%; Score 228; DB 3; Length 46;
Best Local Similarity 95.6%; Pred. No. 1.1e-12;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 52 HMKNSNQIKILGNQSGFLTKGPSKLNDRADSRSLMDQGNPLII 96
|||||
Db 2 HMKNSNQIKILGNQSGFLTKGPSKLNDRADSRSLMDQGNCPPLI 46

RESULT 83

US-09-100-409A-34
Sequence 34, Application US/09100409A
Patent No. 6090388

GENERAL INFORMATION:
APPLICANT: Wang, Chang YI

TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
PREVENTION AND TREATMENT OF HIV INFECTION AND

TITLE OF INVENTION: IMMUNE DISORDERS
NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 Park Avenue
CITY: New York

STATE: NY
COUNTRY: USA

ZIP: 10154-0054

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version
SOFTWARE: #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,409A

FILING DATE:
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME:

REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 1151-4154

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800

TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:
LENGTH: 81 amino acids

TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: peptide
US-09-100-409A-34

Query Match 17.3%; Score 228; DB 3; Length 81;
Best Local Similarity 95.6%; Pred. No. 2.4e-12;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 52 HMKNSNQIKILGNQSGFLTKGPSKLNDRADSRSLMDQGNPLII 96
|||||
Db 37 HMKNSNQIKILGNQSGFLTKGPSKLNDRADSRSLMDQGNCPPLI 81

RESULT 84

US-08-630-172-7

Sequence 7, Application US/08630172
Patent No. 6060054

GENERAL INFORMATION:
APPLICANT: Staez, Uwe

TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
LYMPHOCYTE VETO

NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor

CITY: Denver
STATE: Colorado

COUNTRY: U.S.
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/630,172
FILING DATE:

CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:

LENGTH: 154 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-630-172-7

Query Match 16.1%; Score 212; DB 3; Length 154;
Best Local Similarity 36.2%; Pred. No. 1.3e-10;
Matches 54; Conservative 18; Mismatches 19; Indels 58; Gaps 5;

QY 144 ESP-----GGSSPSVQCRS-----PRGNI-----163
|||
Db 6 QSPHCTTVPGASVNTICSTSGGLRGIVLPOLGPQPDIIYEDGVPTTDRFRGRIDS 65

QY 164 --QGKTLVSQLELDQSGTWC-----TVLQNKVFEKI-DIVPRASA 205
|||||
Db 66 GSDNLTITMRLQLSDTGTTCQALTEVNVYGSGLVLTVERQSGMWRCSAPPASRA 125

QY 206 LPAPPTGSALPDPTASALPDPPASALP 234.
|||||
Db 126 LPAPPTGSALPDPTASALPDPPASALP 154

RESULT 85
US-09-375-419-7
Sequence 7, Application US/09375419
Patent No. 6264950

GENERAL INFORMATION:
APPLICANT: Staez, Uwe

TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
LYMPHOCYTE VETO

NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor

CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/375,419
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,172
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 154 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-375-419-7

Query Match 16.1%; Score 212; DB 3; Length 154;
Best Local Similarity 36.2%; Pred. No. 1.3e-10;
Matches 54; Conservative 18; Mismatches 19; Indels 58; Gaps 5;

QY 144 ESP-----PGSSPVSQCRS-----PRKNT-----163
DB 6 QSPHCTIVPQASVAVITCTSGSLAGIYLPQAGPQDITTYEDGVPTDTRFRGRIDFS 65
QY 164 --QGKTLVSQLELQDSGTWTC-----TYLQNKVYEFKI-DIVPRASA 205
DB 66 GSDNLTITMRLQLSPGTGYTQAITENVVYSGSLVLVTEEGSQGMHRCSDAPPRASA 125

QY 206 LPAPPTGSALPDPQTASALPDPAPASALP 234
DB 126 LPAPPTGSALPDPQTASALPDPAPASALP 154

RESULT 86
US-09-100-409A-2
Sequence 2, Application US/09100409A
Patent No. 6090388
GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: PEPTIDE COMPOSITION FOR PREVENTION AND TREATMENT OF HIV INFECTION AND
TITLE OF INVENTION: IMMUNE DISORDERS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS: 64
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10154-0054
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,409A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:

REFERENCE/DOCKET NUMBER: 1151-4154
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-100-409A-2

Query Match 16.0%; Score 211; DB 3; Length 40;
Best Local Similarity 97.5%; Pred. No. 2.6e-11;
Matches 39; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 52 HMKNSNQIKILGNQGSFLTGPSTLNDRADSRSLMDQGN 91
DB 1 HMKNSNQIKILGNQGSFLTGPSTLNDRADSRSLMDQGN 40

RESULT 87
US-09-100-409A-4
Sequence 4, Application US/09100409A
Patent No. 6090388
GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: PEPTIDE COMPOSITION FOR PREVENTION AND TREATMENT OF HIV INFECTION AND
TITLE OF INVENTION: IMMUNE DISORDERS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS: 64
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10154-0054
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,409A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 1151-4154
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-100-409A-4

Query Match 16.0%; Score 211; DB 3; Length 42;
Best Local Similarity 97.5%; Pred. No. 2.7e-11;
Matches 39; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 52 HMKNSNQIKILGNQGSFLTGPSTLNDRADSRSLMDQGN 91
DB 2 HMKNSNQIKILGNQGSFLTGPSTLNDRADSRSLMDQGN 41

RESULT 88

US-09-100-409A-32
; Sequence 32, Application US/09100409A
; Patent No. 6090388
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang YI
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
; TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND
; TITLE OF INVENTION: IMMUNE DISORDERS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,409A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 1151-4154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 77 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-100-409A-32

Query Match 16.0%; Score 211; DB 3; Length 77;
Best Local Similarity 97.5%; Pred. No. 6.1e-11;
Matches 39; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 52 HMKNSNQIKLNGSGSLTKGSPKLNDRADSRSLMDQGN 91
DB 37 HMKWNQIKILNGSGSLTKGSPKLNDRADSRSLMDQGN 76

RESULT 89
5510256-5
; Patent No. 5510256
; APPLICANT: KIRSCHNER, RICHARD J.;MOTT, JOHN E.;ECKENRODE,
; FRANCES M.;BRUNNER, DAVID P.
; TITLE OF INVENTION: ELIMINATING INTERNAL INITIATION OF
; SOLUBLE CD4 GENE
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/365,995
; FILING DATE: 29-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 13,828
; FILING DATE: 02-FEB-1993
; APPLICATION NUMBER: 562,861
; FILING DATE: 06-AUG-1990
; SEQ ID NO:5:
; LENGTH: 41
5510256-5

Query Match 15.2%; Score 200; DB 6; Length 41;
Best Local Similarity 97.6%; Pred. No. 2.3e-10;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 95 IIKNLIKEDSPYICEVEDEKKEVOLVGLTANSDFHLLQ 135
DB 1 IIKNKKIEDSPYICEVEDEKKEVOLVGLTANSDFHLLQ 41

RESULT 90
US-08-332-562A-85
; Sequence 85, Application US/08332562A
; Patent No. 5985599
; GENERAL INFORMATION:
; APPLICANT: MCKENZIE, Ian F.C.
; APPLICANT: HOGARTH, Mark P.
; APPLICANT: HIBBS, Margaret L.
; APPLICANT: SCOTT, Bernadette M.
; APPLICANT: BONADONNA, Lisa
; APPLICANT: HULETT, Mark D.
; TITLE OF INVENTION: FC RECEPTOR FOR IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,562A
; FILING DATE: 31-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/896,457
; FILING DATE: 27-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 54270/119/GRHA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 81 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-332-562A-85

Query Match 12.8%; Score 168.5; DB 2; Length 81;
Best Local Similarity 50.7%; Pred. No. 2.6e-07;
Matches 36; Conservative 11; Mismatches 23; Indels 1; Gaps 1;

QY 123 FGLTANSDFHLLQGGSLTTLTLES-PPGSSPVQCRSPRKNIQGGTLLSVSQLELQDSGT 181
DB 1 FKVTFSPTGSLTQGGSLTTLTLDNSKVSNPTECKRHKKGKVSQSVLSMSNLRVQDSDP 60

QY 182 WTCVTVLQDNKK 192
DB 61 WNCVTVLQDNKK 71

RESULT 91
US-08-332-562A-86
; Sequence 86, Application US/08332562A
; Patent No. 5985599
; GENERAL INFORMATION:
; APPLICANT: MCKENZIE, Ian F.C.

APPLICANT: HOGARTH, Mark P.
APPLICANT: HIBBS, Margaret L.
APPLICANT: SCOTT, Bernadette M.
APPLICANT: BONADONNA, Lisa
APPLICANT: HULETT, Mark D.
TITLE OF INVENTION: PC RECEPTOR FOR IMMUNOGLOBULIN
NUMBER OF SEQUENCES: 136
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/332,562A
FILING DATE: 31-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/896,457
FILING DATE: 27-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 54270/119/GRHA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 86:
SEQUENCE CHARACTERISTICS:
LENGTH: 81 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-332-562A-86

Query Match 12.7%; Score 167.5; DB 2; Length 81;
Best Local Similarity 44.9%; Pred. No. 3.2e-07;
Matches 35; Conservative 10; Mismatches 32; Indels 1; Gaps 1;

QY 123 FGLTANSDTHLLQGSILTLTLES-PPGSSPSVOCRSRPGKNIQCGKTLVSQLELDPSGT 181
DB 1 FRYTFNPGTRLLQGSILTLTLDSPKVSDDPIECKHSSNIVKSKAFSTHSLAIDPSGI 60

QY 182 WTCVTLONOKKVEFKIDI 199
DB 61 WNCVTVLNOKKGSFDMKL 78

RESULT 92
US-09-100-409A-11
Sequence 11, Application US/09100409A
Patent No. 6090388
GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
PREVENTION AND TREATMENT OF HIV INFECTION AND
TITLE OF INVENTION: IMMUNE DISORDERS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10154-0054
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version
SOFTWARE: #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,409A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 1151-4154
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 34 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-100-409A-11

Query Match 12.7%; Score 167; DB 3; Length 34;
Best Local Similarity 97.0%; Pred. No. 1.1e-07;
Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 64 NQGSFLTGPSTKNDRADSRSLMDQGNFPLII 96
DB 2 NQGSFLTGPSTKNDRADSRSLMDQGNCPILII 34

RESULT 93
US-09-100-409A-33
Sequence 33, Application US/09100409A
Patent No. 6090388
GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
PREVENTION AND TREATMENT OF HIV INFECTION AND
TITLE OF INVENTION: IMMUNE DISORDERS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10154-0054
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version
SOFTWARE: #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,409A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 1151-4154
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 69 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-09-100-409A-33

Query Match 12.7%; Score 167; DB 3; Length 69;
Best Local Similarity 97.0%; Pred. No. 2.8e-07;
Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 64 NQGSFLTKGPSKLNDRADSRSLMDQGNFLII 96
Db 37 NQGSFLTKGPSKLNDRADSRSLMDQGNFLII 69

RESULT 94

US-09-100-409A-19
; Sequence 19, Application US/09100409A
; Patent No. 6090388
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
; TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND
; TITLE OF INVENTION: IMMUNE DISORDERS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,409A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 1151-4154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 29 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-100-409A-19

Query Match 11.8%; Score 156; DB 3; Length 29;
Best Local Similarity 100.0%; Pred. No. 7.6e-07;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 NQGSFLTKGPSKLNDRADSRSLMDQGNF 92
Db 1 NQGSFLTKGPSKLNDRADSRSLMDQGNF 29

RESULT 95
US-09-100-414B-92
; Sequence 92, Application US/09100414B
; Patent No. 6025468
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: NOVEL LHRH PEPTIDE
; TITLE OF INVENTION: IMMUNOGENS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.

STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA

ZIP: 10154-0054
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC Windows
SOFTWARE: Word 97

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,414B
FILING DATE: 20-JUNE-1998
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maria H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4157
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 92:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-100-414B-92

Query Match 11.5%; Score 151.5; DB 3; Length 47;
Best Local Similarity 66.0%; Pred. No. 3.5e-06;
Matches 33; Conservative 5; Mismatches 5; Indels 7; Gaps 2;

QY 44 SQKSIQFHWNKSNQIKILG--NQGSFLTKGPSKLNDRADSRSLMDQGN 91
Db 2 SEIKGVYH-----KIEGIGCNQGSFLTKGPSKLNDRADSRSLMDQGN 46

RESULT 96
US-09-303-323-92
; Sequence 92, Application US/09303323
; Patent No. 6228987
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: NOVEL LHRH PEPTIDE
; TITLE OF INVENTION: IMMUNOGENS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC Windows
; SOFTWARE: Word 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/303,323
; FILING DATE: 30-APR-1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/100,414
; FILING DATE: 20-JUNE-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4157
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849

Search completed: August 3, 2004, 13:17:01
 Job time : 9.59249 secs

NAME: Maria H. Lin
 REGISTRATION NUMBER: 29,323
 REFERENCE/DOCKET NUMBER: 1151-4157
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-758-4800
 TELEFAX: 212-751-6849
 INFORMATION FOR SEQ ID NO: 91:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 30 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-09-100-414B-91

Query Match 11.4%; Score 150; DB 3; Length 30;
 Best Local Similarity 100.0%; Pred. No. 2.6e-06;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 NQGSFLTKGPKSLNDRADSRSLWDQGN 91
 DB 2 NQGSFLTKGPKSLNDRADSRSLWDQGN 29

RESULT 100

US-09-100-409A-5
 Sequence 5, Application US/09100409A
 Patent No. 6090388

GENERAL INFORMATION:

APPLICANT: Wang, Chang Yi

TITLE OF INVENTION: PEPTIDE COMPOSITION FOR

TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND

TITLE OF INVENTION: IMMUNE DISORDERS

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10154-0054

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/100,409A

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME:

REGISTRATION NUMBER:

REFERENCE/DOCKET NUMBER: 1151-4154

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-758-4800

TELEFAX: 212-751-6849

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 30 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-100-409A-5

Query Match 11.4%; Score 150; DB 3; Length 30;
 Best Local Similarity 100.0%; Pred. No. 2.6e-06;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 NQGSFLTKGPKSLNDRADSRSLWDQGN 91
 DB 2 NQGSFLTKGPKSLNDRADSRSLWDQGN 29

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 13:01:34 ; Search time 13.331 Seconds
(without alignments)
1754.300 Million cell updates/sec

Title: SEQ5
Perfect score: 2335
Sequence: 1 MNRGVPRHLLVLIQLALP.....VISFLIGLGIVACVLRTR 453

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 125 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
- 6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2046	88.0	458	3	US-08-466-368-4
2	2043.5	87.9	630	4	US-08-472-888A-6
3	2040	87.7	462	2	US-08-417-495-5
4	2040	87.7	462	2	US-08-284-391B-5
5	2040	87.7	462	3	US-09-218-950-5
6	2040	87.7	462	5	PCT-US92-01785-5
7	2040	87.7	462	5	PCT-US95-00454-5
8	2038.5	87.7	575	2	US-08-417-495-4
9	2038.5	87.7	575	2	US-08-284-391B-4
10	2038.5	87.7	575	3	US-09-218-950-4
11	2038.5	87.7	575	5	PCT-US92-01785-4
12	2038.5	87.7	575	5	PCT-US95-00454-4
13	2038	87.7	457	4	US-08-328-500-9
14	2037	87.6	398	3	US-08-284-391B-29
15	2037	87.6	398	3	US-09-218-950-29
16	2037	87.6	532	2	US-08-417-495-6
17	2037	87.6	532	2	US-08-284-391B-6
18	2037	87.6	532	3	US-09-218-950-6
19	2037	87.6	532	5	PCT-US92-01785-6
20	2037	87.6	532	5	PCT-US95-00454-6
21	2032	87.4	458	3	US-09-039-555B-15
22	2024	87.1	458	4	US-09-517-605-3
23	2017	86.8	402	1	US-08-236-311-1
24	2017	86.8	402	3	US-08-457-918-1
25	2006	86.3	458	6	5223394-7
26	2001	86.1	394	3	US-08-466-368-2
27	2001	86.1	394	4	US-08-328-500-2

28	1951	83.9	394	6	5223418-2	Patent No. 5223418
29	1909	82.1	433	2	US-08-867-149-1	Sequence 1, Appli
30	1909	82.1	433	2	US-08-808-374-1	Sequence 1, Appli
31	1904	81.9	433	3	US-09-100-409A-1	Sequence 1, Appli
32	1904	81.9	434	1	US-08-236-311-4	Sequence 4, Appli
33	1904	81.9	434	3	US-08-457-918-4	Sequence 4, Appli
34	1712	73.6	433	6	5171838-13	Patent No. 5171838
35	1368	58.8	318	6	5223394-11	Patent No. 5223394
36	1363	58.6	295	6	5223394-9	Patent No. 5223394
37	1050.5	45.2	310	3	US-08-477-460B-6	Sequence 6, Appli
38	1050.5	45.2	310	3	US-08-379-516-6	Sequence 6, Appli
39	1050.5	45.2	310	3	US-09-329-916-6	Sequence 6, Appli
40	1050.5	45.2	310	3	US-08-485-372A-6	Sequence 6, Appli
41	1050.5	45.2	310	4	US-09-409-006A-6	Sequence 6, Appli
42	1050.5	45.2	310	4	US-08-484-681-6	Sequence 6, Appli
43	1050.5	45.2	310	5	PCT-US93-07422-6	Sequence 6, Appli
44	1046	45.0	530	3	US-08-477-460B-4	Sequence 4, Appli
45	1046	45.0	530	3	US-08-379-516-4	Sequence 4, Appli
46	1046	45.0	530	3	US-09-329-916-4	Sequence 4, Appli
47	1046	45.0	530	3	US-08-485-372A-4	Sequence 4, Appli
48	1046	45.0	530	4	US-09-409-006A-4	Sequence 4, Appli
49	1046	45.0	530	4	US-08-484-681-4	Sequence 4, Appli
50	1046	45.0	530	5	PCT-US93-07422-4	Sequence 4, Appli
51	1045	44.9	432	3	US-08-477-460B-2	Sequence 2, Appli
52	1045	44.9	432	3	US-08-379-516-2	Sequence 2, Appli
53	1045	44.9	432	3	US-09-329-916-2	Sequence 2, Appli
54	1045	44.9	432	3	US-08-485-372A-2	Sequence 2, Appli
55	1045	44.9	432	4	US-09-409-006A-2	Sequence 2, Appli
56	1045	44.9	432	4	US-08-484-681-2	Sequence 2, Appli
57	1045	44.9	432	5	PCT-US93-07422-2	Sequence 2, Appli
58	1041	44.8	203	2	US-08-284-391B-31	Sequence 31, Appli
59	1041	44.8	203	2	US-09-218-950-31	Sequence 31, Appli
60	999	43.0	457	1	US-08-416-478A-8	Sequence 8, Appli
61	999	43.0	457	2	US-08-474-988B-8	Sequence 8, Appli
62	999	43.0	457	2	US-08-394-442B-8	Sequence 8, Appli
63	912.5	39.2	410	3	US-08-630-172-17	Sequence 17, Appli
64	912.5	39.2	410	3	US-09-375-419-17	Sequence 17, Appli
65	911	39.2	903	1	US-08-021-601-12	Sequence 12, Appli
66	911	39.2	903	1	US-08-082-849B-12	Sequence 12, Appli
67	911	39.2	903	5	PCT-US94-01624-12	Sequence 12, Appli
68	716	30.8	138	3	US-08-630-172-1	Sequence 1, Appli
69	716	30.8	138	3	US-09-375-419-1	Sequence 1, Appli
70	603	25.9	376	2	US-08-558-269-10	Sequence 10, Appli
71	603	25.9	376	2	US-09-410-882-10	Sequence 10, Appli
72	589.5	25.4	383	2	US-08-558-269-6	Sequence 6, Appli
73	589.5	25.4	383	3	US-09-410-882-6	Sequence 6, Appli
74	494	21.2	94	4	US-08-470-998-2	Sequence 2, Appli
75	494	21.2	94	4	US-08-466-368-5	Sequence 5, Appli
76	494	21.2	94	4	US-08-328-500-10	Sequence 10, Appli
77	406	17.5	80	2	US-08-332-562A-84	Sequence 84, Appli
78	392	16.9	74	4	US-09-050-861B-8	Sequence 8, Appli
79	391	16.8	83	2	US-08-332-562A-87	Sequence 87, Appli
80	314	13.5	61	3	US-09-100-409A-20	Sequence 20, Appli
81	288	12.4	58	2	US-08-284-391B-35	Sequence 35, Appli
82	288	12.4	58	3	US-09-218-950-35	Sequence 35, Appli
83	228	9.8	46	3	US-09-100-409A-10	Sequence 10, Appli
84	228	9.8	46	3	US-09-100-409A-34	Sequence 34, Appli
85	211	9.1	40	3	US-09-100-409A-2	Sequence 2, Appli
86	211	9.1	40	3	US-09-100-409A-4	Sequence 4, Appli
87	211	9.1	77	3	US-09-100-409A-32	Sequence 32, Appli
88	200	8.6	41	6	5510256-5	Patent No. 5510256
89	184.5	7.9	154	3	US-08-630-172-7	Sequence 7, Appli
90	184.5	7.9	154	3	US-09-375-419-7	Sequence 7, Appli
91	174.5	7.5	81	2	US-08-332-562A-85	Sequence 85, Appli
92	173.5	7.5	81	2	US-08-332-562A-86	Sequence 86, Appli
93	167	7.2	34	3	US-09-100-409A-11	Sequence 11, Appli
94	167	7.2	69	3	US-09-100-409A-33	Sequence 33, Appli
95	156	6.7	29	3	US-09-100-409A-19	Sequence 19, Appli
96	155	6.7	735	5	PCT-US93-00031-13	Sequence 13, Appli
97	155	6.7	736	5	PCT-US93-00031-15	Sequence 15, Appli
98	155	6.7	739	4	US-08-482-073-6	Sequence 6, Appli
99	155	6.7	739	5	PCT-US93-00031-9	Sequence 9, Appli
100	154	6.6	662	1	US-08-261-304-7	Sequence 7, Appli

101 151.5 6.5 47 3 US-09-100-414B-92 Sequence 92, Appl
102 151.5 6.5 47 3 US-09-303-323-92 Sequence 92, Appl
103 151.5 6.5 47 4 US-09-770-014-92 Sequence 92, Appl
104 150.5 6.5 828 1 US-08-261-304-2 Sequence 2, Appl
105 150 6.5 28 3 US-09-100-409A-3 Sequence 3, Appl
106 150 6.5 30 3 US-09-100-414B-91 Sequence 5, Appl
107 150 6.5 30 3 US-09-100-409A-5 Sequence 91, Appl
108 150 6.5 30 3 US-09-303-323-91 Sequence 91, Appl
109 150 6.5 30 4 US-09-770-014-91 Sequence 91, Appl
110 150 6.5 47 3 US-09-100-414B-93 Sequence 93, Appl
111 150 6.5 47 3 US-09-100-414B-94 Sequence 94, Appl
112 150 6.5 47 3 US-09-100-409A-60 Sequence 60, Appl
113 150 6.5 47 3 US-09-303-323-93 Sequence 93, Appl
114 150 6.5 47 3 US-09-303-323-94 Sequence 94, Appl
115 150 6.5 47 4 US-09-770-014-93 Sequence 93, Appl
116 150 6.5 47 4 US-09-770-014-94 Sequence 94, Appl
117 150 6.5 50 3 US-09-100-409A-37 Sequence 37, Appl
118 144 6.2 64 3 US-09-100-409A-35 Sequence 35, Appl
119 143 6.2 647 4 US-09-009-490A-91 Sequence 5, Appl
120 143 6.2 647 4 US-08-482-073-5 Sequence 11, Appl
121 143 6.2 647 5 PCT-US93-00031-11 Patent No. 5171838
122 139 6.0 26 5 PCT-US93-00031-23 Patent No. 5171838
123 138 5.9 647 5 PCT-US93-00031-23 Patent No. 5171838
124 137.5 5.9 642 1 US-08-217-299-1 Sequence 1, Appl
125 137.5 5.9 698 2 US-08-602-725-36 Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-08-466-368-4
Sequence 4, Application US/08466368
Patent No. 6093539

GENERAL INFORMATION:
APPLICANT: Maddon, Paul J.
APPLICANT: Littman, Dan R.
APPLICANT: Chess, Leonard
APPLICANT: Axel, Richard
APPLICANT: Weiss, Robin
APPLICANT: McDougal, J. S.
TITLE OF INVENTION: DNA ENCODING THE T CELL SURFACE PROTEIN
TITLE OF INVENTION: T4 AND USE OF FRAGMENTS OF T4 IN THE TREATMENT OF AIDS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,368
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 24577-EI-B/JPW/AKC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 458 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-466-368-4
Query Match 88.0%; Score 2046; DB 3; Length 458;
Best Local Similarity 89.6%; Pred. No. 6.2e-163;
Matches 407; Conservative 2; Mismatches 15; Indels 30; Gaps 2;

QY 1 MNRGVFRRLLVQLALPAATQGNKVVLLGKGGDTVELTCTASOKKSIOFHMKNNSQIK 60
Db 1 MNRGVFRRLLVQLALPAATQGNKVVLLGKGGDTVELTCTASOKKSIOFHMKNNSQIK 60
QY 61 ILNQGSLFTKGSFKNDRASSRRSLMDQNFPLIIRNLTIEDSDYTCVEEDQKEVQL 120
Db 61 ILNQGSLFTKGSFKNDRASSRRSLMDQNFPLIIRNLTIEDSDYTCVEEDQKEVQL 120
QY 121 LVGGLTANSTHLLQGSLLTLESPGSSPSVQCSPRGKNIQGGKTLISVSLQELDSC 180
Db 121 LVGGLTANSTHLLQGSLLTLESPGSSPSVQCSPRGKNIQGGKTLISVSLQELDSC 180
QY 181 TWCTVQLNQKVEFKIDIVLAFQKASSIVYKKEGBOVEFSPPLAFVTEKLTGSGELMW 240
Db 181 TWCTVQLNQKVEFKIDIVLAFQKASSIVYKKEGBOVEFSPPLAFVTEKLTGSGELMW 240
QY 241 QABRASSSKSWITFDLKNKEVSVKRYTQPKLQMGKKLPLHLTLPQALPOYAGSGNITLA 300
Db 241 QABRASSSKSWITFDLKNKEVSVKRYTQPKLQMGKKLPLHLTLPQALPOYAGSGNITLA 300
QY 301 LEAKTGKLEHENVLVNMRATQLOKNTLCEYMGPTSPKLMSTLKENKEAKVSKREKPYWV 360
Db 301 LEAKTGKLEHENVLVNMRATQLOKNTLCEYMGPTSPKLMSTLKENKEAKVSKREKPYWV 360
QY 361 LNEBAGMOCCLSDSQVLLSNIKVLPTWSTFVHPRASALPAPPGSALPDQOTASALP 420
Db 361 LNEBAGMOCCLSDSQVLLSNIKVLPTWSTFVHPRASALPAPPGSALPDQOTASALP 420
QY 421 DPAASALPAAALVIFSLGLGLGV-ACVLAR 453
Db 421 DPAASALPAAALVIFSLGLGLGV-ACVLAR 453
QY 453 397 -----MALIVLGGVAGLLFTIGIFFCVRCRR 425
Db 453 397 -----MALIVLGGVAGLLFTIGIFFCVRCRR 425

RESULT 2
US-08-472-888A-6
Sequence 6, Application US/08472888A
Patent No. 6613746
GENERAL INFORMATION:
APPLICANT: Walz, Brian
APPLICANT: Seed, Brian
TITLE OF INVENTION: AGP-ANTIBODY FUSION PROTEINS
TITLE OF INVENTION: AND RELATED MOLECULES AND METHODS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,888A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/618,314
FILING DATE: 23-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/258001
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-472-888A-6

Query Match 87.9%; Score 2043.5; DB 4; Length 630;
Best Local Similarity 95.0%; Pred. No. 1,66-162;
Matches 401; Conservative 2; Mismatches 12; Indels 7; Gaps 1;

```
QY 1 MNRGVPRHLLVLVQLALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHMKNSNOIK 60
DB 1 MNRGVPRHLLVLVQLALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHMKNSNOIK 60
QY 61 ILNGGSLFTGKPSKLNDRADSRSLMDQGNPFLIIKNLKIEDSDTYICEVEDQKEEYQL 120
DB 61 ILNGGSLFTGKPSKLNDRADSRSLMDQGNPFLIIKNLKIEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSDDTHLQGGSLTTLTLESPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDDTHLQGGSLTTLTLESPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLOQOKKVEFKIDIVLAFQKASSIYKKEGQVEFSPLAFIVEKLTGSGELMW 240
DB 181 TWTCVTLOQOKKVEFKIDIVLAFQKASSIYKKEGQVEFSPLAFIVEKLTGSGELMW 240
QY 241 QAERASSSSSWITFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOALPOYAGSGNLTIA 300
DB 241 QAERASSSSSWITFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOALPOYAGSGNLTIA 300
QY 301 LEAKTGKGLHGVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKAKYSKREKPYWV 360
DB 301 LEAKTGKGLHGVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKAKYSKREKPYWV 360
QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVHPRASALPAPTGSALPDPOQTASALP 420
DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVHPRASALPAPTGSALPDPOQTASALP 420
QY 421 DP 422
DB 421 DP 422
QY 414 CP 415
DB 414 CP 415
```

RESULT 3
US-08-417-495-5
Sequence 5, Application US/08417495
Patent No. 5843728
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Redirection of Cellular Immunity by Chimerae
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/417,495
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/203,866
FILING DATE:
APPLICATION NUMBER: US/07/847,566
FILING DATE:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-417-495-5

Query Match 87.7%; Score 2040; DB 2; Length 462;
Best Local Similarity 91.4%; Pred. No. 2e-162;
Matches 405; Conservative 1; Mismatches 9; Indels 28; Gaps 2;

```
QY 1 MNRGVPRHLLVLVQLALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHMKNSNOIK 60
DB 1 MNRGVPRHLLVLVQLALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHMKNSNOIK 60
QY 61 ILNGGSLFTGKPSKLNDRADSRSLMDQGNPFLIIKNLKIEDSDTYICEVEDQKEEYQL 120
DB 61 ILNGGSLFTGKPSKLNDRADSRSLMDQGNPFLIIKNLKIEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSDDTHLQGGSLTTLTLESPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDDTHLQGGSLTTLTLESPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLOQOKKVEFKIDIVLAFQKASSIYKKEGQVEFSPLAFIVEKLTGSGELMW 240
DB 181 TWTCVTLOQOKKVEFKIDIVLAFQKASSIYKKEGQVEFSPLAFIVEKLTGSGELMW 240
QY 241 QAERASSSSSWITFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOALPOYAGSGNLTIA 300
DB 241 QAERASSSSSWITFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOALPOYAGSGNLTIA 300
QY 301 LEAKTGKGLHGVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKAKYSKREKPYWV 360
DB 301 LEAKTGKGLHGVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKAKYSKREKPYWV 360
QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVHPRASALPAPTGSALPDPOQTASALP 420
DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVHPRASALPAPTGSALPDPOQTASALP 420
QY 421 DP 422
DB 421 DP 422
QY 400 -----LCYILDALIFLYGIVL 415
DB 400 -----LCYILDALIFLYGIVL 415
```

RESULT 4
US-08-284-391B-5
Sequence 5, Application US/08284391B
Patent No. 5851828
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS

NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391B
FILING DATE: 02-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-284-391B-5

Query Match 87.7%; Score 2040; DB 2; Length 462;
Best Local Similarity 91.4%; Pred. No. 2e-162;
Matches 405; Conservative 1; Mismatches 9; Indels 28; Gaps 2;
QY 1 MNRGVPFRHLVLLVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHWKNSNOIK 60
DB 1 MNRGVPFRHLVLLVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHWKNSNOIK 60
QY 61 ILNGQSFLLTGKSPKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGQSFLLTGKSPKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGSLLTLTLESPGSSPSVQCRSPRKNIOGGKTLVSQLELDGSG 180
DB 121 LVFGLTANSDTHLLQGSLLTLTLESPGSSPSVQCRSPRKNIOGGKTLVSQLELDGSG 180
QY 121 LVFGLTANSDTHLLQGSLLTLTLESPGSSPSVQCRSPRKNIOGGKTLVSQLELDGSG 180
DB 121 LVFGLTANSDTHLLQGSLLTLTLESPGSSPSVQCRSPRKNIOGGKTLVSQLELDGSG 180
QY 181 TWCTCVLONOKKVEFKIDIVLAFOKASIIYKKEGGEVSEFLAFVEKLTGSGGLMW 240
DB 181 TWCTCVLONOKKVEFKIDIVLAFOKASIIYKKEGGEVSEFLAFVEKLTGSGGLMW 240
QY 241 QAEBAASSKSWITFDLKNKEVSVKRVTDPRKLQMGKLLPLHLTPQALPYAGSGNLTIA 300
DB 241 QAEBAASSKSWITFDLKNKEVSVKRVTDPRKLQMGKLLPLHLTPQALPYAGSGNLTIA 300
QY 301 LEAATGKLGHEVNVLMVRATOLQKNTCEVWGPSPKLMSTLKENKEAVSKKEKPVW 360
DB 301 LEAATGKLGHEVNVLMVRATOLQKNTCEVWGPSPKLMSTLKENKEAVSKKEKPVW 360
QY 361 LNPAGMWOCLLSDSGVLLLESNTKVLPTWSTVHPASALPAPPTGSALPDPQTALALP 420
DB 361 LNPAGMWOCLLSDSGVLLLESNTKVLPTWSTVHPASALPAPPTGSALPDPQTALALP 420
QY 361 LNPAGMWOCLLSDSGVLLLESNTKVLPTWSTVHPASALPAPPTGSALPDPQTALALP 420
DB 361 LNPAGMWOCLLSDSGVLLLESNTKVLPTWSTVHPASALPAPPTGSALPDPQTALALP 420

QY 421 DPPASALPAAVAVISFLGLGL 443
DB 400 -----LCVTLDAILFLVGLVL 415

RESULT 5
US-09-218-950-5
Sequence 5, Application US/09218950
Patent No. 6284240
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/218,950
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-218-950-5

Query Match 87.7%; Score 2040; DB 3; Length 462;
Best Local Similarity 91.4%; Pred. No. 2e-162;
Matches 405; Conservative 1; Mismatches 9; Indels 28; Gaps 2;
QY 1 MNRGVPFRHLVLLVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHWKNSNOIK 60
DB 1 MNRGVPFRHLVLLVQLALPPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHWKNSNOIK 60
QY 61 ILNGQSFLLTGKSPKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGQSFLLTGKSPKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGSLLTLTLESPGSSPSVQCRSPRKNIOGGKTLVSQLELDGSG 180
DB 121 LVFGLTANSDTHLLQGSLLTLTLESPGSSPSVQCRSPRKNIOGGKTLVSQLELDGSG 180

Db 121 LVFGTANSDTHLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
QY 181 TWTCVTLOKQKVEFKIDIVLAFQKASSIYVKKEGEQVEFSPPLAFTVEKLTSGGELMW 240
Db 181 TWTCVTLOKQKVEFKIDIVLAFQKASSIYVKKEGEQVEFSPPLAFTVEKLTSGGELMW 240
QY 241 QAERASSSSKSWITFPDLKKEVSVKRVTPDKLQMGKULPLHLTLPOALPOYAGSGLTLA 300
Db 241 QAERASSSSKSWITFPDLKKEVSVKRVTPDKLQMGKULPLHLTLPOALPOYAGSGLTLA 300
QY 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPYWV 360
Db 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPYWV 360
QY 361 LNPEAGMWOCILSDSGQVLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPOTASALP 420
Db 361 LNPEAGMWOCILSDSGQVLESNIKVLPTWSTPVH-----ADPQ----- 399
QY 421 DPPASALPALAVISFLGLGL 443
Db 400 -----LCYILDALFLYGIYL 415
RESULT 6
PCT-US92-01785-5
; Sequence 5, Application PC/TUS9201785
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation
; TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor
; TITLE OF INVENTION: Chimeras
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/01785
; FILING DATE: 19920306
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/665,961
; FILING DATE: March 7, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/119002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US92-01785-5
Query Match 87.7%; Score 2040; DB 5; Length 462;
Best Local Similarity 91.4%; Pred No. 2e-162;
Matches 405; Conservative 1; Mismatches 9; Indels 28; Gaps 2;
QY 1 NNRGVFRRLLLVQLALPAAATGKGVVVGKGGDTVELTCTASQKSIQFMKNSNDIK 60
Db 1 NNRGVFRRLLLVQLALPAAATGKGVVVGKGGDTVELTCTASQKSIQFMKNSNDIK 60

QY 61 ILNGSFLTKGPSKLANDRADSRRLMDQNPPLIKNLKIEDSDTYICEVEDQKEEYOL 120
Db 61 ILNGSFLTKGPSKLANDRADSRRLMDQNPPLIKNLKIEDSDTYICEVEDQKEEYOL 120
QY 121 LVFGTANSDTHLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
Db 121 LVFGTANSDTHLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
QY 181 TWTCVTLOKQKVEFKIDIVLAFQKASSIYVKKEGEQVEFSPPLAFTVEKLTSGGELMW 240
Db 181 TWTCVTLOKQKVEFKIDIVLAFQKASSIYVKKEGEQVEFSPPLAFTVEKLTSGGELMW 240
QY 241 QAERASSSSKSWITFPDLKKEVSVKRVTPDKLQMGKULPLHLTLPOALPOYAGSGLTLA 300
Db 241 QAERASSSSKSWITFPDLKKEVSVKRVTPDKLQMGKULPLHLTLPOALPOYAGSGLTLA 300
QY 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPYWV 360
Db 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPYWV 360
QY 361 LNPEAGMWOCILSDSGQVLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPOTASALP 420
Db 361 LNPEAGMWOCILSDSGQVLESNIKVLPTWSTPVH-----ADPQ----- 399
QY 421 DPPASALPALAVISFLGLGL 443
Db 400 -----LCYILDALFLYGIYL 415
RESULT 7
PCT-US95-00454-5
; Sequence 5, Application PC/TUS9500454
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian et al.
; TITLE OF INVENTION: Targeted Cytolysis of HIV-Infected
; TITLE OF INVENTION: Cells by Chimeric CD4 Receptor-
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00454
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/847,566
; FILING DATE: March 6, 1992
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/665,961
; FILING DATE: March 7, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/247001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:

LENGTH: 462 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-00454-5

Query Match 87.7%; Score 2040; DB 5; Length 462;
 Best Local Similarity 91.4%; Pred. No. 2e-162;
 Matches 405; Conservative 1; Mismatches 9; Indels 28; Gaps 2;

```

QY 1 MNRGVPRHLLVQLALLPAATQGNKVLGKGGDTVELTCTASQKKSIOFHWKNSQIK 60
DB 1 MNRGVPRHLLVQLALLPAATQGNKVLGKGGDTVELTCTASQKKSIOFHWKNSQIK 60
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIKKLTIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIKKLTIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTCVLQONKKVEFKIDIVLAPQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
DB 181 TWCTCVLQONKKVEFKIDIVLAPQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
QY 241 QAEBASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLA 300
DB 241 QAEBASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLA 300
QY 301 LEAKTGKLGHOENVLVVWRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPVW 360
DB 301 LEAKTGKLGHOENVLVVWRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPVW 360
QY 361 LNPEAGMWQCLSDSGVLLIESNIKVLPTWSTVHADPKCLYLLDGLFTYGVITLALYL 420
DB 361 LNPEAGMWQCLSDSGVLLIESNIKVLPTWSTVHADPKCLYLLDGLFTYGVITLALYL 420
QY 421 DPBAASALPALAVISFLGLGL 443
DB 400 -----LCYTLDAITFLYGIYL 415

```

RESULT 8
 US-08-417-495-4
 Sequence 4, Application US/08417495
 Patent No. 5843728

GENERAL INFORMATION:
 APPLICANT: Seed, Brian et al.
 TITLE OF INVENTION: Redirection of Cellular Immunity by Chimerae
 NUMBER OF SEQUENCES: 27
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804

COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 COMPUTER: IBM PS/2 Model 502 or 555X
 OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
 SOFTWARE: Wordperfect (Version 5.0)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/417,495

CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/203,866
 FILING DATE:
 APPLICATION NUMBER: US/07/847,566
 FILING DATE:

APPLICATION NUMBER: 07/665,961
 FILING DATE: March 7, 1991
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Clark, Paul T.
 REGISTRATION NUMBER: 30,162
 REFERENCE/DOCKET NUMBER: 00786/119002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 542-5070
 TELEFAX: (617) 542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 575 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-417-495-4

Query Match 87.7%; Score 2038.5; DB 2; Length 575;
 Best Local Similarity 92.1%; Pred. No. 3.7e-162;
 Matches 406; Conservative 4; Mismatches 8; Indels 23; Gaps 3;

```

QY 1 MNRGVPRHLLVQLALLPAATQGNKVLGKGGDTVELTCTASQKKSIOFHWKNSQIK 60
DB 1 MNRGVPRHLLVQLALLPAATQGNKVLGKGGDTVELTCTASQKKSIOFHWKNSQIK 60
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIKKLTIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIKKLTIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTCVLQONKKVEFKIDIVLAPQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
DB 181 TWCTCVLQONKKVEFKIDIVLAPQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
QY 241 QAEBASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLA 300
DB 241 QAEBASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLA 300
QY 301 LEAKTGKLGHOENVLVVWRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPVW 360
DB 301 LEAKTGKLGHOENVLVVWRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPVW 360
QY 361 LNPEAGMWQCLSDSGVLLIESNIKVLPTWSTVHADPKCLYLLDGLFTYGVITLALYL 420
DB 361 LNPEAGMWQCLSDSGVLLIESNIKVLPTWSTVHADPKCLYLLDGLFTYGVITLALYL 420
QY 402 PAPETGSALPDPOATASALPDP 422
DB 421 RAKFRSA-----ETAAVLQDP 437

```

RESULT 9
 US-08-284-391B-4

Sequence 4, Application US/08284391B
 Patent No. 5851828
 GENERAL INFORMATION:
 APPLICANT: Seed, Brian
 APPLICANT: Banapour, Babak
 APPLICANT: Romeo, Charles
 APPLICANT: Kolanus, Waldemar
 TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
 NUMBER OF SEQUENCES: 53
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Clark & Elbing LLP
 STREET: 176 Federal Street
 CITY: Boston
 STATE: MA

COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391B
FILING DATE: 02-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Eibling, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-284-391B-4

Query Match 87.7% Score 2038.5; DB 2; Length 575;
Best Local Similarity 92.1% Pred. No. 3.7e-162;

Matches 406; Conservative 4; Mismatches 8; Indels 23; Gaps 3;

QY 1 MNRGVFPHLLLVLTQALLPAATOGNKVVLGKKGDTVELTCTASOKSIOFHMKNSNOIK 60
DB 1 MNRGVFPHLLLVLTQALLPAATOGNKVVLGKKGDTVELTCTASOKSIOFHMKNSNOIK 60
QY 61 ILNGGSLFTKGPSKLNDRADSRSLMDQGNPPLIKLKIEDSTYICEVEDOKEEVOL 120
DB 61 ILNGGSLFTKGPSKLNDRADSRSLMDQGNPPLIKLKIEDSTYICEVEDOKEEVOL 120
QY 121 LVFGLTANSDBTHLLOGQSILTLTLESPGSSPVOCRSFRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDBTHLLOGQSILTLTLESPGSSPVOCRSFRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTVTLONOKKVEFKIDIVLAFQKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240
DB 181 TWCTVTLONOKKVEFKIDIVLAFQKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240
QY 241 QAEKASSSSKSWITFDLKNKEVSKRYTQPKLQMGKPLHLTLQALPQVYAGSNLTILA 300
DB 241 QAEKASSSSKSWITFDLKNKEVSKRYTQPKLQMGKPLHLTLQALPQVYAGSNLTILA 300
QY 301 LEAKGKGLHGEVNLVVMRATOLQKULTCVWGPTSPKMLSLKLNKEKAVSKREKPVW 360
DB 301 LEAKGKGLHGEVNLVVMRATOLQKULTCVWGPTSPKMLSLKLNKEKAVSKREKPVW 360
QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVH--PR-----ASAL 401
DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVH--PR-----ASAL 401
QY 402 PAPPTGSALPDQPTASALPDP 422
DB 422 PAPPTGSALPDQPTASALPDP 422
QY 421 RAKFRSA---ETANLQDP 437
DB 437 RAKFRSA---ETANLQDP 437

RESULT 10

US-09-218-950-4
Sequence 4, Application US/09218950
Patent No. 6284240
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Bibling LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/218,950
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195,395
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/847,566
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Eibling, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-218-950-4

Query Match 87.7% Score 2038.5; DB 3; Length 575;
Best Local Similarity 92.1% Pred. No. 3.7e-162;

Matches 406; Conservative 4; Mismatches 8; Indels 23; Gaps 3;

QY 1 MNRGVFPHLLLVLTQALLPAATOGNKVVLGKKGDTVELTCTASOKSIOFHMKNSNOIK 60
DB 1 MNRGVFPHLLLVLTQALLPAATOGNKVVLGKKGDTVELTCTASOKSIOFHMKNSNOIK 60
QY 61 ILNGGSLFTKGPSKLNDRADSRSLMDQGNPPLIKLKIEDSTYICEVEDOKEEVOL 120
DB 61 ILNGGSLFTKGPSKLNDRADSRSLMDQGNPPLIKLKIEDSTYICEVEDOKEEVOL 120
QY 121 LVFGLTANSDBTHLLOGQSILTLTLESPGSSPVOCRSFRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDBTHLLOGQSILTLTLESPGSSPVOCRSFRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTVTLONOKKVEFKIDIVLAFQKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240
DB 181 TWCTVTLONOKKVEFKIDIVLAFQKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240

Qy 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPKLQMGKKLPLHLTLPOLPYAGSGNLTLLA 300
Db 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPKLQMGKKLPLHLTLPOLPYAGSGNLTLLA 300
Qy 301 LEATGKGLHOEVNLYVMRATQLOKNTCEVWGPTSPKMLSLKLENKAKVSKREKPVW 360
Db 301 LEATGKGLHOEVNLYVMRATQLOKNTCEVWGPTSPKMLSLKLENKAKVSKREKPVW 360
Qy 361 LNPEAGMWOCCLSDSGVLESNIKVLPTWSTPVH--PR-----ASAL 401
Db 361 LNPEAGMWOCCLSDSGVLESNIKVLPTWSTPVHADPKLCYLLDGLFTYGVITLTYL 420
Qy 402 PAPPTGSALPDPQTASALPDP 422
Db 421 RAKFSRSA-----ETANLQDP 437

RESULT 11
PCT-US92-01785-4
Sequence 4, Application PC/TUS9201785
GENERAL INFORMATION:
APPLICANT: The General Hospital Corporation
TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/01785
FILING DATE: 19920306
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-01785-4

Query Match 87.7%; Score 2038.5; DB 5; Length 575;
Best Local Similarity 92.1%; Pred. No. 3.7e-162;
Matches 406; Conservative 4; Mismatches 8; Indels 23; Gaps 3;
Qy 1 MNRGVFPHLLVLTALPAPATGKNKVLGKGDVTELTCTASOKSIOFHMKNNOIK 60
Db 1 MNRGVFPHLLVLTALPAPATGKNKVLGKGDVTELTCTASOKSIOFHMKNNOIK 60
Qy 61 ILNGSGFLTKGSPSLNDRADSRSLMDQGNPPLIKKLTEDSDTYICEVEDQKEEYOL 120
Db 61 ILNGSGFLTKGSPSLNDRADSRSLMDQGNPPLIKKLTEDSDTYICEVEDQKEEYOL 120
Qy 121 LVFGLTANSDFHLLQGQSLTTLTLESPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180

Db 121 LVFGLTANSDFHLLQGQSLTTLTLESPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180
Qy 181 TWCTVLOKOKKVEFKDIDIVLAFORASSIVYKKEGQVEFSPPLATVTKLTGSSGLMW 240
Db 181 TWCTVLOKOKKVEFKDIDIVLAFORASSIVYKKEGQVEFSPPLATVTKLTGSSGLMW 240
Qy 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPKLQMGKKLPLHLTLPOLPYAGSGNLTLLA 300
Db 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPKLQMGKKLPLHLTLPOLPYAGSGNLTLLA 300
Qy 301 LEATGKGLHOEVNLYVMRATQLOKNTCEVWGPTSPKMLSLKLENKAKVSKREKPVW 360
Db 301 LEATGKGLHOEVNLYVMRATQLOKNTCEVWGPTSPKMLSLKLENKAKVSKREKPVW 360
Qy 361 LNPEAGMWOCCLSDSGVLESNIKVLPTWSTPVH--PR-----ASAL 401
Db 361 LNPEAGMWOCCLSDSGVLESNIKVLPTWSTPVHADPKLCYLLDGLFTYGVITLTYL 420
Qy 402 PAPPTGSALPDPQTASALPDP 422
Db 421 RAKFSRSA-----ETANLQDP 437

RESULT 12
PCT-US95-00454-4
Sequence 4, Application PC/TUS9500454
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Targeted Cytolysis of HIV-Infected
TITLE OF INVENTION: Cells by Chimeric CD4 Receptor-
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00454
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-00454-4

```

Query Match      87.7%: Score 2038.5; DB 5; Length 575;
Best Local Similarity 92.1%: Pred. No. 3.7e-162;
Matches 406; Conservative 4; Mismatches 8; Indels 23; Gaps 3;

QY 1 NMRGVPFRLHLVLQALLPAATQGNKVVLGKKGDVETLTASQKKSIOFHMKNSNQIK 60
DB 1 NMRGVPFRLHLVLQALLPAATQGNKVVLGKKGDVETLTASQKKSIOFHMKNSNQIK 60
QY 61 IINGGSEFLTKGPSKLNDRADSRSLMDQGNPFLIIKNIKIEDSDTYICEVDQKEEYQL 120
DB 61 IINGGSEFLTKGPSKLNDRADSRSLMDQGNPFLIIKNIKIEDSDTYICEVDQKEEYQL 120
QY 121 LVFGLTANSDBTHLQGGSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDBTHLQGGSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 121 LVFGLTANSDBTHLQGGSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDBTHLQGGSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNOQKVEFKIDIVLAFQKASSIYVKKGEQVEFSPPLAFTVEKLTGSGELMW 240
DB 181 TWTCTVLQNOQKVEFKIDIVLAFQKASSIYVKKGEQVEFSPPLAFTVEKLTGSGELMW 240
QY 241 QAEKASSSKSWITFDLKNKEVSVKRVTPDKLQMGKKLPLHLTLPQALPOYAGSGLTLA 300
DB 241 QAEKASSSKSWITFDLKNKEVSVKRVTPDKLQMGKKLPLHLTLPQALPOYAGSGLTLA 300
QY 301 LEAKTGKLEHGVNVLVVRATOLQKNLTCEWGPSPKMLSLKLEKKAQSKREKPVAV 360
DB 301 LEAKTGKLEHGVNVLVVRATOLQKNLTCEWGPSPKMLSLKLEKKAQSKREKPVAV 360
QY 361 INPEAGMOCCLSDSGVLLSNIKVLPTWSTPVH--PR-----ASAL 401
DB 361 INPEAGMOCCLSDSGVLLSNIKVLPTWSTPVHADPKLQMGKKLPLHLTLPQALPOYAGSGLTLA 420
QY 402 PAPPGSALPDPTQASALPD 422
DB 421 RAKFSRSA---ETANLQDP 437

RESULT 13
US-08-328-500-9
; Sequence 9, Application US/08328500
; Patent No. 6673896
; GENERAL INFORMATION:
; APPLICANT: Maddon, Paul J.
; APPLICANT: Axel, Richard W.
; APPLICANT: Sweet, Richard W.
; APPLICANT: Arthos, James
; TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,500
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/24577-CY
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 457 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-500-9

Query Match      87.7%: Score 2038; DB 4; Length 457;
Best Local Similarity 89.4%: Pred. No. 2.9e-162;
Matches 406; Conservative 2; Mismatches 16; Indels 30; Gaps 2;

QY 1 NMRGVPFRLHLVLQALLPAATQGNKVVLGKKGDVETLTASQKKSIOFHMKNSNQIK 60
DB 1 NMRGVPFRLHLVLQALLPAATQGNKVVLGKKGDVETLTASQKKSIOFHMKNSNQIK 60
QY 61 IINGGSEFLTKGPSKLNDRADSRSLMDQGNPFLIIKNIKIEDSDTYICEVDQKEEYQL 120
DB 61 IINGGSEFLTKGPSKLNDRADSRSLMDQGNPFLIIKNIKIEDSDTYICEVDQKEEYQL 120
QY 121 LVFGLTANSDBTHLQGGSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDBTHLQGGSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 121 LVFGLTANSDBTHLQGGSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDBTHLQGGSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNOQKVEFKIDIVLAFQKASSIYVKKGEQVEFSPPLAFTVEKLTGSGELMW 240
DB 181 TWTCTVLQNOQKVEFKIDIVLAFQKASSIYVKKGEQVEFSPPLAFTVEKLTGSGELMW 240
QY 241 QAEKASSSKSWITFDLKNKEVSVKRVTPDKLQMGKKLPLHLTLPQALPOYAGSGLTLA 300
DB 241 QAEKASSSKSWITFDLKNKEVSVKRVTPDKLQMGKKLPLHLTLPQALPOYAGSGLTLA 300
QY 301 LEAKTGKLEHGVNVLVVRATOLQKNLTCEWGPSPKMLSLKLEKKAQSKREKPVAV 360
DB 301 LEAKTGKLEHGVNVLVVRATOLQKNLTCEWGPSPKMLSLKLEKKAQSKREKPVAV 360
QY 361 INPEAGMOCCLSDSGVLLSNIKVLPTWSTPVHPRASALPAPPTGASALPDPTQASALP 420
DB 361 INPEAGMOCCLSDSGVLLSNIKVLPTWSTPVHPRASALPAPPTGASALPDPTQASALP 420
QY 421 DPPASALPALAVISFLIGLGLV-ACVLAATR 453
DB 397 -----MALIVLGVAGVGLLFIGLIGFVCRGHR 425

RESULT 14
US-08-284-391B-29
; Sequence 29, Application US/08284391B
; Patent No. 5851828
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Banapour, Babak
; APPLICANT: Romeo, Charles
; APPLICANT: Kolanus, Waldemar
; TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
; TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,391B
; FILING DATE: 02-AUG-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/195,395

```

FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 398 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-284-391B-29

Query Match 87.6%; Score 2037; DB 2; Length 398;
Best Local Similarity 100.0%; Pred. No. 2.8e-162;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRLHLVQLALLPAATQGNKRVYLGKKGDVTELTCTASOKSIQPHWKNNSQIK 60
DB 1 MNRGVPFRLHLVQLALLPAATQGNKRVYLGKKGDVTELTCTASOKSIQPHWKNNSQIK 60
QY 61 ILNGQSFLLTKGSPSKLNDRAISRSLMDQGNFPIIKLIKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGQSFLLTKGSPSKLNDRAISRSLMDQGNFPIIKLIKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGGSLTLLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLQGGSLTLLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLOKQKVEFKIDIVVLAFOKASSIVYKKEGEVQFSPLATVEKLTGSGGLMW 240
DB 181 TWTCVTLOKQKVEFKIDIVVLAFOKASSIVYKKEGEVQFSPLATVEKLTGSGGLMW 240
QY 241 QAEARSSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPQALPYAGSGNLTLLA 300
DB 241 QAEARSSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPQALPYAGSGNLTLLA 300
QY 301 LEAKTGKLGHOEVNLVVMRATQKNIJTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
DB 301 LEAKTGKLGHOEVNLVVMRATQKNIJTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
QY 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTPVH 395
DB 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTPVH 395

RESULT 15

US-09-218-950-29
Sequence 29, Application US/09218950

PATENT NO. 6284240
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waidemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESS: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA

ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/218,950
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 398 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-218-950-29

Query Match 87.6%; Score 2037; DB 3; Length 398;
Best Local Similarity 100.0%; Pred. No. 2.8e-162;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRLHLVQLALLPAATQGNKRVYLGKKGDVTELTCTASOKSIQPHWKNNSQIK 60
DB 1 MNRGVPFRLHLVQLALLPAATQGNKRVYLGKKGDVTELTCTASOKSIQPHWKNNSQIK 60
QY 61 ILNGQSFLLTKGSPSKLNDRAISRSLMDQGNFPIIKLIKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGQSFLLTKGSPSKLNDRAISRSLMDQGNFPIIKLIKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGGSLTLLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLQGGSLTLLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLOKQKVEFKIDIVVLAFOKASSIVYKKEGEVQFSPLATVEKLTGSGGLMW 240
DB 181 TWTCVTLOKQKVEFKIDIVVLAFOKASSIVYKKEGEVQFSPLATVEKLTGSGGLMW 240
QY 241 QAEARSSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPQALPYAGSGNLTLLA 300
DB 241 QAEARSSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPQALPYAGSGNLTLLA 300
QY 301 LEAKTGKLGHOEVNLVVMRATQKNIJTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
DB 301 LEAKTGKLGHOEVNLVVMRATQKNIJTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
QY 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTPVH 395
DB 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTPVH 395

RESULT 16

US-08-417-495-6
Sequence 6, Application US/08417495
Patent No. 5843728

```
/ GENERAL INFORMATION:
/ APPLICANT: Seed, Brian et al.
/ TITLE OF INVENTION: Redirection of Cellular Immunity by
/ TITLE OF INVENTION: Receptor
/ NUMBER OF SEQUENCES: 27
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSEE: Fish & Richardson
/ STREET: 225 Franklin Street
/ CITY: Boston
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02110-2804
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ COMPUTER: IBM PS/2 Model 502 or 55SX
/ OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
/ SOFTWARE: Wordperfect (Version 5.0)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/417,495
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/203,866
/ FILING DATE:
/ APPLICATION NUMBER: US/07/847,566
/ FILING DATE:
/ APPLICATION NUMBER: 07/665,961
/ FILING DATE: March 7, 1991
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Clark, Paul T.
/ REGISTRATION NUMBER: 30,162
/ REFERENCE/DOCKET NUMBER: 00786/119002
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 542-5070
/ TELEFAX: (617) 542-8906
/ TELEX: 200154
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 532 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-417-495-6

Query Match      87.6%; Score 2037; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 4,4e-162;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLLVLTQALLPAAATGKNVVLGKKGDTVELTCTAQQKSIQPHMKNNSQIK 60
DB 1 MNRGVPFRHLLVLTQALLPAAATGKNVVLGKKGDTVELTCTAQQKSIQPHMKNNSQIK 60
QY 61 ILNGQSFLLTGKPSKLNDRADSRSLMDQGNFLLIKNLKIETSDTYICEVEDQKEEVQL 120
DB 61 ILNGQSFLLTGKPSKLNDRADSRSLMDQGNFLLIKNLKIETSDTYICEVEDQKEEVQL 120
QY 121 LVFGLTANSDFHLLQGQSLTLLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
DB 121 LVFGLTANSDFHLLQGQSLTLLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
QY 181 TWTCTVLTQONOKKVEKIDIVLAFOKASSIYKKEGEQEVRSPLAFTVEKLTGSGELMW 240
DB 181 TWTCTVLTQONOKKVEKIDIVLAFOKASSIYKKEGEQEVRSPLAFTVEKLTGSGELMW 240
QY 241 QAERASSSSKSWITFDLKNKKEVSVKRVTDQPKLQMGKLLPLHLTLPOALPOYAGSNTLTLA 300
DB 241 QAERASSSSKSWITFDLKNKKEVSVKRVTDQPKLQMGKLLPLHLTLPOALPOYAGSNTLTLA 300
QY 301 LEAKTGKGLHQQEVNVLVMBRATOLQKNLTCEVWGPTSPKLMLSIKENKEAKVSKREKPVAV 360
DB 301 LEAKTGKGLHQQEVNVLVMBRATOLQKNLTCEVWGPTSPKLMLSIKENKEAKVSKREKPVAV 360
```

```
QY 361 LNPEAGMQCILSDSGVLTLESNIKVLPTMSTPVA 395
DB 361 LNPEAGMQCILSDSGVLTLESNIKVLPTMSTPVA 395

RESULT 17
US-08-284-391B-6
Sequence 6, Application US/08284391B
Patent No. 5851828
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391B
FILING DATE: 02-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-284-391B-6

Query Match      87.6%; Score 2037; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 4,4e-162;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLLVLTQALLPAAATGKNVVLGKKGDTVELTCTAQQKSIQPHMKNNSQIK 60
DB 1 MNRGVPFRHLLVLTQALLPAAATGKNVVLGKKGDTVELTCTAQQKSIQPHMKNNSQIK 60
QY 61 ILNGQSFLLTGKPSKLNDRADSRSLMDQGNFLLIKNLKIETSDTYICEVEDQKEEVQL 120
DB 61 ILNGQSFLLTGKPSKLNDRADSRSLMDQGNFLLIKNLKIETSDTYICEVEDQKEEVQL 120
QY 121 LVFGLTANSDFHLLQGQSLTLLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
DB 121 LVFGLTANSDFHLLQGQSLTLLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
```

Qy 181 TWCTVLONOKKVEFKIDIVLAFQKASSIVYKKEGQVFSPLAFTVEKLTGSGELMW 240
Db 181 TWCTVLONOKKVEFKIDIVLAFQKASSIVYKKEGQVFSPLAFTVEKLTGSGELMW 240
Qy 241 QAEKASSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPOALPOYAGSGNLTLLA 300
Db 241 QAEKASSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPOALPOYAGSGNLTLLA 300
Qy 301 LEAKTGKHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPYWV 360
Db 301 LEAKTGKHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPYWV 360
Qy 361 LNPEAGMWQCLSDSGQVLLSENIKVLPTWSTPVH 395
Db 361 LNPEAGMWQCLSDSGQVLLSENIKVLPTWSTPVH 395

RESULT 18
US-09-218-950-6
Sequence 6, Application US/09218950
Patent No. 6284240
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Pasteo for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/218, 950
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-218-950-6

Query Match 87.6%; Score 2037; DB 3; Length 532;
Best Local Similarity 100.0%; Pred. No. 4,4e-162;

Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNRGVPFRHLVLVQLALPAATQGNKVYLGKKDFTVELCTASQKSIQFHMKNNSQIK 60
Db 1 MNRGVPFRHLVLVQLALPAATQGNKVYLGKKDFTVELCTASQKSIQFHMKNNSQIK 60
Qy 61 ILNQGSLFTKGPSKLNDRASRRSLMDQGNFPLIKNLTIEDSDTYICEVEDQKEVQL 120
Db 61 ILNQGSLFTKGPSKLNDRASRRSLMDQGNFPLIKNLTIEDSDTYICEVEDQKEVQL 120
Qy 121 LVFGLTANSTHLLQGSQSLTLTLESPGSSPSVQCSPPKRNIOGGKTLTSSQLTLDSC 180
Db 121 LVFGLTANSTHLLQGSQSLTLTLESPGSSPSVQCSPPKRNIOGGKTLTSSQLTLDSC 180
Qy 181 TWCTVLONOKKVEFKIDIVLAFQKASSIVYKKEGQVFSPLAFTVEKLTGSGELMW 240
Db 181 TWCTVLONOKKVEFKIDIVLAFQKASSIVYKKEGQVFSPLAFTVEKLTGSGELMW 240
Qy 241 QAEKASSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPOALPOYAGSGNLTLLA 300
Db 241 QAEKASSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPOALPOYAGSGNLTLLA 300
Qy 301 LEAKTGKHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPYWV 360
Db 301 LEAKTGKHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPYWV 360
Qy 361 LNPEAGMWQCLSDSGQVLLSENIKVLPTWSTPVH 395
Db 361 LNPEAGMWQCLSDSGQVLLSENIKVLPTWSTPVH 395

RESULT 19
PCT-US92-01785-6
Sequence 6, Application PC/TUS9201785
GENERAL INFORMATION:
APPLICANT: The General Hospital Corporation
TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/01785
FILING DATE: 19920306
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-01785-6

Query Match 87.6%; Score 2037; DB 5; Length 532;
Best Local Similarity 100.0%; Pred. No. 4.4e-162;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLILVLQALIPAAATQGNKVLGKKGDVLTCTASQKSIQFMKXSNQIK 60
DB 1 MNRGVPFRHLILVLQALIPAAATQGNKVLGKKGDVLTCTASQKSIQFMKXSNQIK 60
QY 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNPFLIIKNLKIEDSTYICEVEDQKEEYQL 120
DB 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNPFLIIKNLKIEDSTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSDBTHLQGGSLTLTLESPGSSPSVQCRSPRGNIIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDBTHLQGGSLTLTLESPGSSPSVQCRSPRGNIIQGGKTLVSQLELDQSG 180
QY 181 TWCTVLOKQKVEFKIDIVLAFOKASSIYKKGEQVEFSFPLAFVTEKLTSGSELW 240
DB 181 TWCTVLOKQKVEFKIDIVLAFOKASSIYKKGEQVEFSFPLAFVTEKLTSGSELW 240
QY 241 QAEKSSSKSWITFDLKNKEVSVKRVTDPKLQMGKULPLHLTLPQALPOYAGSNLTIA 300
DB 241 QAEKSSSKSWITFDLKNKEVSVKRVTDPKLQMGKULPLHLTLPQALPOYAGSNLTIA 300
QY 301 LEAKTGKLGHOEVNLVWMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKYSKREKPYW 360
DB 301 LEAKTGKLGHOEVNLVWMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKYSKREKPYW 360
QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVH 395
DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVH 395

RESULT 20
PCT-US95-00454-6
; Sequence 6, Application PC/TUS9500454
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian et al.
; TITLE OF INVENTION: Targeted Cytolysis of HIV-Infected
; TITLE OF INVENTION: Cells by Chimeric CD4 Receptor-
; TITLE OF INVENTION: Bearing Cells
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 558X
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00454
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/847,566
; FILING DATE: March 6, 1992
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/665,961
; FILING DATE: March 7, 1991
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/247001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-00454-6

Query Match 87.6%; Score 2037; DB 5; Length 532;
Best Local Similarity 100.0%; Pred. No. 4.4e-162;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLILVLQALIPAAATQGNKVLGKKGDVLTCTASQKSIQFMKXSNQIK 60
DB 1 MNRGVPFRHLILVLQALIPAAATQGNKVLGKKGDVLTCTASQKSIQFMKXSNQIK 60
QY 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNPFLIIKNLKIEDSTYICEVEDQKEEYQL 120
DB 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNPFLIIKNLKIEDSTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSDBTHLQGGSLTLTLESPGSSPSVQCRSPRGNIIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDBTHLQGGSLTLTLESPGSSPSVQCRSPRGNIIQGGKTLVSQLELDQSG 180
QY 181 TWCTVLOKQKVEFKIDIVLAFOKASSIYKKGEQVEFSFPLAFVTEKLTSGSELW 240
DB 181 TWCTVLOKQKVEFKIDIVLAFOKASSIYKKGEQVEFSFPLAFVTEKLTSGSELW 240
QY 241 QAEKSSSKSWITFDLKNKEVSVKRVTDPKLQMGKULPLHLTLPQALPOYAGSNLTIA 300
DB 241 QAEKSSSKSWITFDLKNKEVSVKRVTDPKLQMGKULPLHLTLPQALPOYAGSNLTIA 300
QY 301 LEAKTGKLGHOEVNLVWMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKYSKREKPYW 360
DB 301 LEAKTGKLGHOEVNLVWMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKYSKREKPYW 360
QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVH 395
DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPVH 395

RESULT 21
US-09-039-555B-15
; Sequence 15, Application US/09039555B
; Patent No. 6033856
; GENERAL INFORMATION:
; APPLICANT: Koerner, Kathrin
; APPLICANT: Mueller, Rolf
; TITLE OF INVENTION: Sadiacek, Hans-Harald
; TITLE OF INVENTION: PROMOTER OF THE CDC25B GENE, ITS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Hardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/039,555B
; FILING DATE: 16-MAR-1998
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 19710643.9
; FILING DATE: 14-MAR-1997

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016779/0131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 458 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-039-555B-15

Query Match      87.4%; Score 2032; DB 3; Length 458;
Best Local Similarity 89.2%; Pred. No. 9,1e-162;
Matches 405; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

QY      1 MNRGVPRHLLLVQLALPAAATGKNVYLGKGDVVELTCTASOKKSIOFHMKNQIK 60
DB      1 MNRGVPRHLLLVQLALPAAATGKKNVYLGKGDVVELTCTASOKKSIOFHMKNQIK 60
QY      61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEEVOL 120
DB      61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEEVOL 120
QY      121 LVFGLTANSDTHLLQGOSLTLTLESPGSSPSVQCSRPGRKNIQGGKTLVSQLELDG 180
DB      121 LVFGLTANSDTHLLQGOSLTLTLESPGSSPSVQCSRPGRKNIQGGKTLVSQLELDG 180
QY      121 LVFGLTANSDTHLLQGOSLTLTLESPGSSPSVQCSRPGRKNIQGGKTLVSQLELDG 180
DB      121 LVFGLTANSDTHLLQGOSLTLTLESPGSSPSVQCSRPGRKNIQGGKTLVSQLELDG 180
QY      181 TWCTVLOKQKVEFKIDIVLAFQKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240
DB      181 TWCTVLOKQKVEFKIDIVLAFQKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240
QY      181 TWCTVLOKQKVEFKIDIVLAFQKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240
DB      181 TWCTVLOKQKVEFKIDIVLAFQKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240
QY      241 QABRASSSKSWITFDLKNKEVSVKRYTODPKLQMGKKLPHLTLPQALPYAGSGNLTIA 300
DB      241 QABRASSSKSWITFDLKNKEVSVKRYTODPKLQMGKKLPHLTLPQALPYAGSGNLTIA 300
QY      301 LEAKTGKLGHEVNLVVMRATOLQKNLTCEVWGPTSPKLMSTLKENKEAVSKREKPVW 360
DB      301 LEAKTGKLGHEVNLVVMRATOLQKNLTCEVWGPTSPKLMSTLKENKEAVSKREKPVW 360
QY      361 LNPBAGMWQCLSDSGOVLESNIKVLPTWSTVPHPRASALPAPPTGSALPDQTASALP 420
DB      361 LNPBAGMWQCLSDSGOVLESNIKVLPTWSTVPHPRASALPAPPTGSALPDQTASALP 420
QY      421 DPPAASALPALAVISFLGLGLGV-ACVLARTR 453
DB      397 -----MALIVLGVAAGLLIFGLGIFCVCRHR 425

RESULT 22
US-09-517-605-3
; Sequence 3, Application US/09517605
; Patent No. 6391567
; GENERAL INFORMATION:
; APPLICANT: Littman, Dan R.
; APPLICANT: Kwon, Douglas S.
; APPLICANT: van Kooyk, Yvette
; APPLICANT: Geiltebeck, Theo
; TITLE OF INVENTION: METHODS OF USING A FACILITATOR OF RETROVIRAL ENTRY INTO
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 1049-1-017
; CURRENT APPLICATION NUMBER: US/09/517, 605
; CURRENT FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 458
; TYPE: PRT
```

```

; ORGANISM: Homo sapiens
; US-09-517-605-3

Query Match      87.1%; Score 2024; DB 4; Length 458;
Best Local Similarity 89.0%; Pred. No. 4,3e-161;
Matches 404; Conservative 2; Mismatches 18; Indels 30; Gaps 2;

QY      1 MNRGVPRHLLLVQLALPAAATGKNVYLGKGDVVELTCTASOKKSIOFHMKNQIK 60
DB      1 MNRGVPRHLLLVQLALPAAATGKKNVYLGKGDVVELTCTASOKKSIOFHMKNQIK 60
QY      61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEEVOL 120
DB      61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEEVOL 120
QY      121 LVFGLTANSDTHLLQGOSLTLTLESPGSSPSVQCSRPGRKNIQGGKTLVSQLELDG 180
DB      121 LVFGLTANSDTHLLQGOSLTLTLESPGSSPSVQCSRPGRKNIQGGKTLVSQLELDG 180
QY      121 LVFGLTANSDTHLLQGOSLTLTLESPGSSPSVQCSRPGRKNIQGGKTLVSQLELDG 180
DB      121 LVFGLTANSDTHLLQGOSLTLTLESPGSSPSVQCSRPGRKNIQGGKTLVSQLELDG 180
QY      181 TWCTVLOKQKVEFKIDIVLAFQKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240
DB      181 TWCTVLOKQKVEFKIDIVLAFQKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240
QY      241 QABRASSSKSWITFDLKNKEVSVKRYTODPKLQMGKKLPHLTLPQALPYAGSGNLTIA 300
DB      241 QABRASSSKSWITFDLKNKEVSVKRYTODPKLQMGKKLPHLTLPQALPYAGSGNLTIA 300
QY      301 LEAKTGKLGHEVNLVVMRATOLQKNLTCEVWGPTSPKLMSTLKENKEAVSKREKPVW 360
DB      301 LEAKTGKLGHEVNLVVMRATOLQKNLTCEVWGPTSPKLMSTLKENKEAVSKREKPVW 360
QY      361 LNPBAGMWQCLSDSGOVLESNIKVLPTWSTVPHPRASALPAPPTGSALPDQTASALP 420
DB      361 LNPBAGMWQCLSDSGOVLESNIKVLPTWSTVPHPRASALPAPPTGSALPDQTASALP 420
QY      421 DPPAASALPALAVISFLGLGLGV-ACVLARTR 453
DB      397 -----MALIVLGVAAGLLIFGLGIFCVCRHR 425

RESULT 23
US-08-236-311-1
; Sequence 1, Application US/08236311
; Patent No. 5565335
; GENERAL INFORMATION:
; APPLICANT: Capon, Daniel J.
; APPLICANT: Gregory, Timothy J.
; TITLE OF INVENTION: Adhesion Variants
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/236,311
; FILING DATE: 02-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/936190
; FILING DATE: 26-AUG-1992
; PRIOR APPLICATION DATA: 07/842777
; APPLICATION NUMBER: 07/842777
; FILING DATE: 18-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/250785
```



```

; FILING DATE: 28-SEP-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/104329
; FILING DATE: 02-OCT-1987
; ATTORNEY/AGENT INFORMATION:
; NAME: Habak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 444P1C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1996
; TELEFAX: 415/952-9881
; TELEEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-236-311-1

```

```

Query Match      86.8%; Score 2017; DB 1; Length 402;
Best Local Similarity 99.7%; Pred. No. 1,4e-160;
Matches 392; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 MNRGVPFRHLVLVQLALPAATQGNKVVLGKKGDTVELTCTASQKSIQFHMKNNOIK 60
   |||||||
DB 1 MNRGVPFRHLVLVQLALPAATQGNKVVLGKKGDTVELTCTASQKSIQFHMKNNOIK 60
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEEYQL 120
   |||||||
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSDTHLLOGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
   |||||||
DB 121 LVFGLTANSDTHLLOGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTCTVLONQKKVEFKIDIVLAFQKASSIYKKEGEVEFSFPLAFTVEKLTSGGELMW 240
   |||||||
DB 181 TWCTCTVLONQKKVEFKIDIVLAFQKASSIYKKEGEVEFSFPLAFTVEKLTSGGELMW 240
QY 241 QAEKRRSSSKSWITPDLKKEVSVKRVTDPPKLOMGKKLPLHLTLPALPOYAGSNNLTLA 300
   |||||||
DB 241 QAEKRRSSSKSWITPDLKKEVSVKRVTDPPKLOMGKKLPLHLTLPALPOYAGSNNLTLA 300
QY 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCGVWGPTSPKMLSLKLNKEAKVSKREKPYVW 360
   |||||||
DB 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCGVWGPTSPKMLSLKLNKEAKVSKREKPYVW 360
QY 361 LNPEAGMMQCLLSDSGQVLLSNIKVLPTWSTP 393
   |||||||
DB 361 LNPEAGMMQCLLSDSGQVLLSNIKVLPTWSTP 393

```

```

RESULT 24
US-08-457-918-1
; Sequence 1, Application US/08457918
; Patent No. 6117655
; GENERAL INFORMATION:
; APPLICANT: Capon, Daniel J.
; APPLICANT: Gregory, Timothy J.
; TITLE OF INVENTION: Adhesion Variance
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,918
; FILING DATE: 1-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/236311
; FILING DATE: 02-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/936190
; FILING DATE: 26-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/842777
; FILING DATE: 18-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/250785
; FILING DATE: 28-SEP-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/104329
; FILING DATE: 02-OCT-1987
; ATTORNEY/AGENT INFORMATION:
; NAME: Kubienc, Jeffrey S.
; REGISTRATION NUMBER: 36,575
; REFERENCE/DOCKET NUMBER: P0444P1C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8228
; TELEFAX: 415/952-9881
; TELEEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-457-918-1

```

```

Query Match      86.8%; Score 2017; DB 3; Length 402;
Best Local Similarity 99.7%; Pred. No. 1,4e-160;
Matches 392; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 MNRGVPFRHLVLVQLALPAATQGNKVVLGKKGDTVELTCTASQKSIQFHMKNNOIK 60
   |||||||
DB 1 MNRGVPFRHLVLVQLALPAATQGNKVVLGKKGDTVELTCTASQKSIQFHMKNNOIK 60
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEEYQL 120
   |||||||
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSDTHLLOGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
   |||||||
DB 121 LVFGLTANSDTHLLOGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTCTVLONQKKVEFKIDIVLAFQKASSIYKKEGEVEFSFPLAFTVEKLTSGGELMW 240
   |||||||
DB 181 TWCTCTVLONQKKVEFKIDIVLAFQKASSIYKKEGEVEFSFPLAFTVEKLTSGGELMW 240
QY 241 QAEKRRSSSKSWITPDLKKEVSVKRVTDPPKLOMGKKLPLHLTLPALPOYAGSNNLTLA 300
   |||||||
DB 241 QAEKRRSSSKSWITPDLKKEVSVKRVTDPPKLOMGKKLPLHLTLPALPOYAGSNNLTLA 300
QY 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCGVWGPTSPKMLSLKLNKEAKVSKREKPYVW 360
   |||||||
DB 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCGVWGPTSPKMLSLKLNKEAKVSKREKPYVW 360
QY 361 LNPEAGMMQCLLSDSGQVLLSNIKVLPTWSTP 393
   |||||||
DB 361 LNPEAGMMQCLLSDSGQVLLSNIKVLPTWSTP 393

```

```

RESULT 25
5223394-7
; Patent No. 5223394
; APPLICANT: WALINER, BARBARA
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULE COMPRISING
; LYMPHOCTE FUNCTION-ASSOCIATED ANTIGEN 3 PHOSPHATIDYLINOSITOL

```

LINKAGE SIGNAL SEQUENCE

NUMBER OF SEQUENCES: 12

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/335,688

FILING DATE: 10-APR-1989

SEQ ID NO: 7

LENGTH: 458

5223394-7

Query Match 86.3%; Score 2006; DB 6; Length 458;

Best Local Similarity 88.3%; Pred. No. 1.4e-159;

Matches 401; Conservative 2; Mismatches 21; Indels 30; Gaps 2;

QY 1 MNRGVPRHLLVLTQALPPATQGNKVVLGKGGDTVELTCTASOKKSIOFHMKNSNQIK 60
 DB 1 MNRGVPRHLLVLTQALPPATQGNKVVLGKGGDTVELTCTASOKKSIOFHMKNSNQIK 60
 QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDYICVEEDQKEEVOL 120
 DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDYICVEEDQKEEVOL 120
 QY 121 LVFGLTANSPTHLLQGSGLTLTLSPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
 DB 121 LVFGLTANSPTHLLQGSGLTLTLSPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
 QY 181 TWCTVLQONQKVEFKIDIVLAFOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
 DB 181 TWCTVLQONQKVEFKIDIVLAFOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
 QY 241 QAEKASSSKSWITFDLKNKEVSVKRVTPDKLQMGKKLPLHLTPOLPOYAGSGNLTLA 300
 DB 241 QAEKASSSKSWITFDLKNKEVSVKRVTPDKLQMGKKLPLHLTPOLPOYAGSGNLTLA 300
 QY 301 LEAKTGKLGHOEVNLVVMRATOLQKNTCEVWGPTSPKMLSLKLENKAKVSKREKPVW 360
 DB 301 LEAKTGKLGHOEVNLVVMRATOLQKNTCEVWGPTSPKMLSLKLENKAKVSKREKPVW 360
 QY 361 LNPEAGMWQCLLSDSGQVLTLESNIKVLPTWSTPVHPRAALPAPPTGALPDPQTASALP 420
 DB 361 LNPEAGMWQCLLSDSGQVLTLESNIKVLPTWSTPVHPRAALPAPPTGALPDPQTASALP 420
 QY 421 DPEASALPALAVISFLGLGLGV-ACVLAATR 453
 DB 397 -----MALIVLGVAAGLLFLGLGIFPCVRCRHR 425

RESULT 26

US-08-466-368-2

Sequence 2, Application US/08466368

Patent No. 6093539

GENERAL INFORMATION:

APPLICANT: Maddon, Paul J.

APPLICANT: Littman, Dan R.

APPLICANT: Chase, Leonard

APPLICANT: Axel, Richard

APPLICANT: Weis, Robin

APPLICANT: McDougal, J. S.

TITLE OF INVENTION: DNA ENCODING THE T CELL SURFACE PROTEIN

TITLE OF INVENTION: T4 AND USE OF FRAGMENTS OF T4 IN THE TREATMENT OF AIDS

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP

STREET: 1185 Avenue of Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,368

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 24577-BI-B/JPW/AKC

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-278-0400

TELEFAX: 212-391-0525

INFORMATION FOR SEQ ID NO: 2

SEQUENCE CHARACTERISTICS:

LENGTH: 394 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-466-368-2

Query Match 86.1%; Score 2001; DB 3; Length 394;

Best Local Similarity 98.7%; Pred. No. 2.9e-159;

Matches 389; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLTQALPPATQGNKVVLGKGGDTVELTCTASOKKSIOFHMKNSNQIK 60
 DB 1 MNRGVPRHLLVLTQALPPATQGNKVVLGKGGDTVELTCTASOKKSIOFHMKNSNQIK 60
 QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDYICVEEDQKEEVOL 120
 DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDYICVEEDQKEEVOL 120
 QY 121 LVFGLTANSPTHLLQGSGLTLTLSPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
 DB 121 LVFGLTANSPTHLLQGSGLTLTLSPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
 QY 181 TWCTVLQONQKVEFKIDIVLAFOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
 DB 181 TWCTVLQONQKVEFKIDIVLAFOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMW 240
 QY 241 QAEKASSSKSWITFDLKNKEVSVKRVTPDKLQMGKKLPLHLTPOLPOYAGSGNLTLA 300
 DB 241 QAEKASSSKSWITFDLKNKEVSVKRVTPDKLQMGKKLPLHLTPOLPOYAGSGNLTLA 300
 QY 301 LEAKTGKLGHOEVNLVVMRATOLQKNTCEVWGPTSPKMLSLKLENKAKVSKREKPVW 360
 DB 301 LEAKTGKLGHOEVNLVVMRATOLQKNTCEVWGPTSPKMLSLKLENKAKVSKREKPVW 360
 QY 361 LNPEAGMWQCLLSDSGQVLTLESNIKVLPTWSTPV 394
 DB 361 LNPEAGMWQCLLSDSGQVLTLESNIKVLPTWSTPV 394

RESULT 27

US-08-328-500-2

Sequence 2, Application US/08328500

Patent No. 6673896

GENERAL INFORMATION:

APPLICANT: Maddon, Paul J.

APPLICANT: Sweet, Richard W.

APPLICANT: Athos, James

APPLICANT: Derivatives of Soluble T-4

TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:

```
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,500
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/24577-CY
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-328-500-2

Query Match      86.1%; Score 2001; DB 4: Length 394;
Best Local Similarity 98.7%; Pred. No. 2,9e-159;
Matches 389; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NMRGVPFRLLVLVQLALLPAATQGNKVLGKKGDVVELTCTASQKSIQFHMKNQIK 60
DB 1 NMRGVPFRLLVLVQLALLPAATQGNKVLGKKGDVVELTCTASQKSIQFHMKNQIK 60

QY 61 ILGNQGSFLTKGPPSKINDRADSRSLMDQGNFPLIIRNKIKEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPPSKINDRADSRSLMDQGNFPLIIRNKIKEDSDTYICEVEDQKEEVOL 120

QY 121 LVFGLTANSDTHLLOQGSLLTLESPPSSPVQCRSPRGKNIQGGKTLISVQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQGSLLTLESPPSSPVQCRSPRGKNIQGGKTLISVQLELDQSG 180

QY 181 TWTCVTLQNKQKVEFKIDIVLAFOKASSIYVKEGEQVEFSPFLAFTVEKLTGSGELMW 240
DB 181 TWTCVTLQNKQKVEFKIDIVLAFOKASSIYVKEGEQVEFSPFLAFTVEKLTGSGELMW 240

QY 241 TWTCVTLQNKQKVEFKIDIVLAFOKASSIYVKEGEQVEFSPFLAFTVEKLTGSGELMW 240
DB 241 TWTCVTLQNKQKVEFKIDIVLAFOKASSIYVKEGEQVEFSPFLAFTVEKLTGSGELMW 240

QY 241 QAEKSSSKSWITPDLKNKEVSVKRVTDPKLQMGKPLHLTLPOLPYAGSGNLTILA 300
DB 241 QAEKSSSKSWITPDLKNKEVSVKRVTDPKLQMGKPLHLTLPOLPYAGSGNLTILA 300

QY 301 LEAKTGKGLHGVNLVWMATOLQKULTCVWGPSPKMLSLKLENKEAKYSKEKPYAV 360
DB 301 LEAKTGKGLHGVNLVWMATOLQKULTCVWGPSPKMLSLKLENKEAKYSKEKPYAV 360

QY 361 LNPEAGMMOCLLSDSGVLLSNIKVLPTWSTPV 394
DB 361 LNPEAGMMOCLLSDSGVLLSNIKVLPTWSTPV 394

RESULT 28
5223418-2
PATENT No. 5223418
APPLICANT: ARCURI, EDWARD J.;BRAMNER, MARY E.;DONOVAN, MARY
J.;GERBER, ROBERT G.;KELLER, JOHN A.
TITLE OF INVENTION: METHOD OF IMPROVING THE YIELD OF
HETEROLOGOUS PROTEINS PRODUCED BY STREPTOMYCES LIVIDANS
NUMBER OF SEQUENCES: 2
CURRENT APPLICATION DATA:
REGISTRATION NUMBER: US/07/569,979
REFERENCE/DOCKET NUMBER: 1151-4145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)751-6849
TELEFAX: (212)751-6849
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-867-149-1

Query Match      83.9%; Score 1951; DB 6: Length 394;
Best Local Similarity 97.5%; Pred. No. 4.4e-155;
Matches 384; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 NMRGVPFRLLVLVQLALLPAATQGNKVLGKKGDVVELTCTASQKSIQFHMKNQIK 60
```

```
DB 1 NMRGVPFRLLVLVQLALLPAATQGNKVLGKKGDVVELTCTASQKSIQFHMKNQIK 60

QY 61 ILGNQGSFLTKGPPSKINDRADSRSLMDQGNFPLIIRNKIKEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPPSKINDRADSRSLMDQGNFPLIIRNKIKEDSDTYICEVEDQKEEVOL 120

QY 121 LVFGLTANSDTHLLOQGSLLTLESPPSSPVQCRSPRGKNIQGGKTLISVQLELDQSG 180
DB 121 LVFGLTANSDTHLLOQGSLLTLESPPSSPVQCRSPRGKNIQGGKTLISVQLELDQSG 180

QY 181 TWTCVTLQNKQKVEFKIDIVLAFOKASSIYVKEGEQVEFSPFLAFTVEKLTGSGELMW 240
DB 181 TWTCVTLQNKQKVEFKIDIVLAFOKASSIYVKEGEQVEFSPFLAFTVEKLTGSGELMW 240

QY 241 QAEKSSSKSWITPDLKNKEVSVKRVTDPKLQMGKPLHLTLPOLPYAGSGNLTILA 300
DB 241 QAEKSSSKSWITPDLKNKEVSVKRVTDPKLQMGKPLHLTLPOLPYAGSGNLTILA 300

QY 301 LEAKTGKGLHGVNLVWMATOLQKULTCVWGPSPKMLSLKLENKEAKYSKEKPYAV 360
DB 301 LEAKTGKGLHGVNLVWMATOLQKULTCVWGPSPKMLSLKLENKEAKYSKEKPYAV 360

QY 361 LNPEAGMMOCLLSDSGVLLSNIKVLPTWSTPV 394
DB 361 LNPEAGMMOCLLSDSGVLLSNIKVLPTWSTPV 394

RESULT 29
US-08-867-149-1
Sequence 1, Application US/08867149
PATENT No. 5912176
GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: Antibody Against a Host Cell
TITLE OF INVENTION: Antigen Complex for Pre- and Post-Exposure
PROTECTION FROM INFECTION BY HIV PRIMARY ISOLATES
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSER: Maria C.H. Lin
CITY: New York
STREET: 345 Park Avenue
STATE: NY
COUNTRY: USA
ZIP: 10154-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version
SOFTWARE: #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,149
FILING DATE: 28-Feb-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/657,149
FILING DATE: 03-June-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maria C.H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4145
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)751-6849
TELEFAX: (212)751-6849
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-867-149-1
```

Query Match 82.1%; Score 1909; DB 2; Length 433;
Best Local Similarity 88.6%; Pred. No. 1.6e-151;
Matches 380; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

QY 26 NKVVLGKGGDTVELTCTASQKSIQPHWKNSNOIKILGNQGSFLTKGPSKLNDRADSRRS 85
DB 1 NKVVLGKGGDTVELTCTASQKSIQPHWKNSNOIKILGNQGSFLTKGPSKLNDRADSRRS 60
QY 86 LMDQGNFPLIIKNIKIEDSDTYICEVEDQKEEYQLVFGLTANSDFHLLQGSLTLTLES 145
DB 61 LMDQGNFPLIIKNIKIEDSDTYICEVEDQKEEYQLVFGLTANSDFHLLQGSLTLTLES 120
QY 146 PGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLOKQKVEFKIDIVLAFQ 205
DB 121 PGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLOKQKVEFKIDIVLAFQ 180
QY 206 KASSIVYKKEGEQVEFSFPLAFVTEKLTGSGELMWQAEARASSKSWITPDKNKEYSVKR 265
DB 181 KASSIVYKKEGEQVEFSFPLAFVTEKLTGSGELMWQAEARASSKSWITPDKNKEYSVKR 240
QY 266 VTQDPKLOMGKKLPPLHLTPQALPOYAGSGNLTALAEKTKLHOEVNLYVMRATOLQKN 325
DB 241 VTQDPKLOMGKKLPPLHLTPQALPOYAGSGNLTALAEKTKLHOEVNLYVMRATOLQKN 300
QY 326 LTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCCLSDSGVLLSNIK 385
DB 301 LTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCCLSDSGVLLSNIK 360
QY 386 VLPTWSTPVHPRASALPAPPTGSALPDPTASALPDPPASALPALAVISFLGLGLGV 445
DB 361 VLPTWSTPVHPRASALPAPPTGSALPDPTASALPDPPASALPALAVISFLGLGLGV 391
QY 446 -ACVLARTR 453
DB 392 FFCVRCRHR 400

RESULT 30

US-08-808-374-1
Sequence 1, Application US/08808374
Patent No. 5961976

GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: Antibody Against a Host Cell
TITLE OF INVENTION: Antigen Complex for Pre- and Post-Exposure
TITLE OF INVENTION: Protection from Infection by HIV Primary Isolates
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Maria C.H. Lin
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10154-0053

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version
SOFTWARE: #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808,374
FILING DATE: 28-Feb-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/657,149

FILING DATE: 03-June-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4145

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)415-8745

TELEFAX: (212)751-6849

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 433 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-808-374-1

Query Match 82.1%; Score 1909; DB 2; Length 433;
Best Local Similarity 88.6%; Pred. No. 1.6e-151;
Matches 380; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

QY 26 NKVVLGKGGDTVELTCTASQKSIQPHWKNSNOIKILGNQGSFLTKGPSKLNDRADSRRS 85
DB 1 NKVVLGKGGDTVELTCTASQKSIQPHWKNSNOIKILGNQGSFLTKGPSKLNDRADSRRS 60
QY 86 LMDQGNFPLIIKNIKIEDSDTYICEVEDQKEEYQLVFGLTANSDFHLLQGSLTLTLES 145
DB 61 LMDQGNFPLIIKNIKIEDSDTYICEVEDQKEEYQLVFGLTANSDFHLLQGSLTLTLES 120
QY 146 PGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLOKQKVEFKIDIVLAFQ 205
DB 121 PGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLOKQKVEFKIDIVLAFQ 180
QY 206 KASSIVYKKEGEQVEFSFPLAFVTEKLTGSGELMWQAEARASSKSWITPDKNKEYSVKR 265
DB 181 KASSIVYKKEGEQVEFSFPLAFVTEKLTGSGELMWQAEARASSKSWITPDKNKEYSVKR 240
QY 266 VTQDPKLOMGKKLPPLHLTPQALPOYAGSGNLTALAEKTKLHOEVNLYVMRATOLQKN 325
DB 241 VTQDPKLOMGKKLPPLHLTPQALPOYAGSGNLTALAEKTKLHOEVNLYVMRATOLQKN 300
QY 326 LTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCCLSDSGVLLSNIK 385
DB 301 LTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCCLSDSGVLLSNIK 360
QY 386 VLPTWSTPVHPRASALPAPPTGSALPDPTASALPDPPASALPALAVISFLGLGLGV 445
DB 361 VLPTWSTPVHPRASALPAPPTGSALPDPTASALPDPPASALPALAVISFLGLGLGV 391
QY 446 -ACVLARTR 453
DB 392 FFCVRCRHR 400

RESULT 31

US-09-100-409A-1
Sequence 1, Application US/09100409A
Patent No. 6090388

GENERAL INFORMATION:

APPLICANT: Wang, Chang Yi

TITLE OF INVENTION: PEPTIDE COMPOSITION FOR

TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND

TITLE OF INVENTION: IMMUNE DISORDERS

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10154-0054

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version

SOFTWARE: #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/100,409A

```
/ FILING DATE:
/ CLASSIFICATION: 514
/ ATTORNEY/AGENT INFORMATION:
/ NAME:
/ REGISTRATION NUMBER:
/ REFERENCE/DOCKET NUMBER: 1151-4154
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 212-758-4800
/ TELEFAX: 212-751-6849
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 433 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
US-09-100-409A-1

Query Match      81.9%; Score 1904; DB 3; Length 433;
Best Local Similarity 88.6%; Pred. No. 4.3e-151;
Matches 379; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

QY 27 KVLGKGGDTVELTCTASQKSIQPHWKNKSNQIKILGNQSFLLTKGPSKLNDRADSRSL 86
DB 2 KVLGKGGDTVELTCTASQKSIQPHWKNKSNQIKILGNQSFLLTKGPSKLNDRADSRSL 61
QY 87 WQGNPPLIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLLQGOSLTLTLSP 146
DB 62 WQGNPPLIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLLQGOSLTLTLSP 121
QY 147 PSSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWCTTVLQNKKEFKIDIVVLAPOK 206
DB 122 PSSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWCTTVLQNKKEFKIDIVVLAPOK 181
QY 207 ASSIYKKEGEVSEFPFLAFTVEKLTGSGELMWAERASSKSWITFDLNKKEVSVKRV 266
DB 182 ASSIYKKEGEVSEFPFLAFTVEKLTGSGELMWAERASSKSWITFDLNKKEVSVKRV 241
QY 267 TODPPLQMGKKLPLHLTLPOALPOYAGSGNLTALAEATGKLHOEVNLVWVRATOLQ 326
DB 242 TODPPLQMGKKLPLHLTLPOALPOYAGSGNLTALAEATGKLHOEVNLVWVRATOLQ 301
QY 327 TCEWVGPTSPKMLSLKLENKEAKYSKREKPYWVNLNPEAGMOCLLSDSGVLLSESNIKV 386
DB 302 TCEWVGPTSPKMLSLKLENKEAKYSKREKPYWVNLNPEAGMOCLLSDSGVLLSESNIKV 361
QY 387 LFTWSTPVHPRASALPAPPTGSALPDPTQASALPDPRASALPALAVISFLGLGLGV- 445
DB 362 LFTWSTPVQP-----MALIVLGVAAGLLFLGLGIF 392
QY 446 ACVLARTR 453
DB 393 FCVRCRHR 400

RESULT 32
US-08-236-311-4
/ Sequence 4, Application US/08236311
/ Patent No. 5565335
/ GENERAL INFORMATION:
/ APPLICANT: Capon, Daniel J.
/ APPLICANT: Gregory, Timothy J.
/ TITLE OF INVENTION: Adhesion Variants
/ NUMBER OF SEQUENCES: 25
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Genentech, Inc.
/ STREET: 460 Point San Bruno Blvd
/ CITY: South San Francisco
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94080
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
/ COMPUTER: IBM PC compatible
```

```
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: patin (Genentech)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/236,311
/ FILING DATE: 02-MAY-1994
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/936190
/ FILING DATE: 26-AUG-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/842777
/ FILING DATE: 18-FEB-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/250785
/ FILING DATE: 28-SEP-1988
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/104329
/ FILING DATE: 02-OCT-1987
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Hasak, Janet E.
/ REGISTRATION NUMBER: 28,616
/ REFERENCE/DOCKET NUMBER: 444PIC2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415/225-1896
/ TELEFAX: 415/952-9881
/ TELEX: 910/371-7168
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 434 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
US-08-236-311-4

Query Match      81.9%; Score 1904; DB 1; Length 434;
Best Local Similarity 99.7%; Pred. No. 4.3e-151;
Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 24 QGNKVVLGKKDDTVELTCTASQKSIQPHWKNKSNQIKILGNQSFLLTKGPSKLNDRASR 83
DB 56 QGNKVVLGKKDDTVELTCTASQKSIQPHWKNKSNQIKILGNQSFLLTKGPSKLNDRASR 115
QY 84 RSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLLQGOSLTLTL 143
DB 116 RSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLLQGOSLTLTL 175
QY 144 ESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWCTTVLQNKKEFKIDIVVLA 203
DB 176 ESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWCTTVLQNKKEFKIDIVVLA 235
QY 204 FOKASSIYKKEGEVSEFPFLAFTVEKLTGSGELMWAERASSKSWITFDLNKKEVSV 263
DB 236 FOKASSIYKKEGEVSEFPFLAFTVEKLTGSGELMWAERASSKSWITFDLNKKEVSV 295
QY 264 KRVTDPPKQMGKKLPLHLTLPOALPOYAGSGNLTALAEATGKLHOEVNLVWVRATOLQ 323
DB 296 KRVTDPPKQMGKKLPLHLTLPOALPOYAGSGNLTALAEATGKLHOEVNLVWVRATOLQ 355
QY 324 KNLTCWVGPTSPKMLSLKLENKEAKYSKREKPYWVNLNPEAGMOCLLSDSGVLLSESN 383
DB 356 KNLTCWVGPTSPKMLSLKLENKEAKYSKREKPYWVNLNPEAGMOCLLSDSGVLLSESN 415
QY 384 IKVLPWSTP 393
DB 416 IKVLPWSTP 425

RESULT 33
US-08-457-918-4
/ Sequence 4, Application US/08457918
/ Patent No. 6117655
/ GENERAL INFORMATION:
/ APPLICANT: Capon, Daniel J.
/ APPLICANT: Gregory, Timothy J.
```

```

/ TITLE OF INVENTION: Adhesion Variants
/ NUMBER OF SEQUENCES: 25
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Genentech, Inc.
/ STREET: 460 Point San Bruno Blvd
/ CITY: South San Francisco
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94080
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: patin (Genentech)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/457,918
/ FILING DATE: 1-JUN-1995
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/236311
/ FILING DATE: 02-MAY-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/936190
/ FILING DATE: 26-AUG-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/842777
/ FILING DATE: 18-FEB-1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/250785
/ FILING DATE: 28-SEP-1988
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 07/104329
/ FILING DATE: 02-OCT-1987
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Kubinec, Jeffrey S.
/ REGISTRATION NUMBER: 36,575
/ REFERENCE/DOCKET NUMBER: P0444PIC3
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415/225-8228
/ TELEFAX: 415/952-9881
/ TELEX: 910/371-7168
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 434 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ US-08-457-918-4

Query Match      81.9%; Score 1904; DB 3; Length 434;
Best Local Similarity 99.7%; Pred. No. 4.3e-151;
Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 24 QGNVYVGGKGDVVELTCTASQKKSIOFHWNKSNQIKILGNQGSFLTGPBKLNDRADSR 83
Db 56 QGNVYVGGKGDVVELTCTASQKKSIOFHWNKSNQIKILGNQGSFLTGPBKLNDRADSR 115
Oy 84 RSLMDQGNFPLIIKILKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLLOQOSLTTL 143
Db 116 RSLMDQGNFPLIIKILKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLLOQOSLTTL 175
Oy 144 ESPGSSPSVOCRSFPRGKNIQGGKTLVSQLELDQSGTWCTVLOKQKVEFKIDIVVLA 203
Db 176 ESPGSSPSVOCRSFPRGKNIQGGKTLVSQLELDQSGTWCTVLOKQKVEFKIDIVVLA 235
Oy 204 FOKASSIYKKEGEVEFSFPLAFVEKLTGSGELMWQABRASSKSWITFDLKNKEVS 263
Db 236 FOKASSIYKKEGEVEFSFPLAFVEKLTGSGELMWQABRASSKSWITFDLKNKEVS 295
Oy 264 KRVTQDPKIQMGKPLHLTLPOALPOYAGSGNLTALAEKTKGLHQBENLVWMRAVDLO 323
Db 296 KRVTQDPKIQMGKPLHLTLPOALPOYAGSGNLTALAEKTKGLHQBENLVWMRAVDLO 355
Oy 324 KNLTCFVWGPTSPKLMLSLKLLENKAVSKREKPVVVLNBPAGMOCCLSDSGVLLSEN 383

```

```

Db 356 KNLTCFVWGPTSPKLMLSLKLLENKAVSKREKPVVVLNBPAGMOCCLSDSGVLLSEN 415
Oy 384 IKVLPWTSTP 393
Db 416 IKVLPWTSTP 425

RESULT 34
5171838-13
/ Patent No. 5171838
/ APPLICANT: CHIBA, YUKINOBU
/ TITLE OF INVENTION: LEU3A BINDING PEPTIDES
/ NUMBER OF SEQUENCES: 24
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/526,921
/ FILING DATE: 22-MAY-1990
/ SEQ ID NO:13:
/ LENGTH: 433
5171838-13

```

```

Query Match      73.6%; Score 1712; DB 6; Length 433;
Best Local Similarity 81.6%; Pred. No. 5e-135;
Matches 355; Conservative 8; Mismatches 28; Indels 44; Gaps 6;

```

```

Oy 27 KVLVGGKGDVVELTCTASQKKSIOFHWNKSNQIKILGNQGSFLTGPBKLNDRADSR- 85
Db 2 KVLVGGKGDVVELTCTASQKKSIOFHWNKSNQIKILGNQGSFLTGPBKLNDRADSR- 61
Oy 86 ----LMDQGNFPLIIKILKIED-SDTYICEVEDQKEEVQLVFGLTANSPTHLLOQOSLT 139
Db 62 NQIKILGNKGSF-LTGKPSKLNDRADSR-----RSEEVQLVFGLTANSPTHLLOQOSLT 114
Oy 140 TLTLESPGSSPSVOCRSFPRGKNIQGGKTLVSQLELDQSGTWCTVLOKQKVEFKIDI 199
Db 115 TLTLESPGSSPSVOCRSFPRGKNIQGGKTLVSQLELDQSGTWCTVLOKQKVEFKIDI 174
Oy 200 VVLAFOKASSIYKKEGEVEFSFPLAFVEKLTGSGELMWQABRASSKSWITFDLKNK 259
Db 175 VVLAFOKASSIYKKEGEVEFSFPLAFVEKLTGSGELMWQABRASSKSWITFDLKNK 234
Oy 260 EVSVKRYTQDPKIQMGKPLHLTLPOALPOYAGSGNLTALAEKTKGLHQBENLVWMRA 319
Db 235 EVSVKRYTQDPKIQMGKPLHLTLPOALPOYAGSGNLTALAEKTKGLHQBENLVWMRA 294
Oy 320 TOLQKNLTCFVWGPTSPKLMLSLKLLENKAVSKREKPVVVLNBPAGMOCCLSDSGOVL 379
Db 295 TOLQKNLTCFVWGPTSPKLMLSLKLLENKAVSKREKPVVVLNBPAGMOCCLSDSGOVL 354
Oy 380 LESNIKVLPTWSTPVHPRASALPAPPTGALPDPTASALPDPPAASALPALAVISFLL 439
Db 355 LESNIKVLPTWSTPVQ-----MALVLGSGVALL 385
Oy 440 GLGIGV-ACVLAATR 453
Db 386 FIGIGIFCVCRHR 400

```

```

RESULT 35
5223394-11
/ Patent No. 5223394
/ APPLICANT: WALLNER, BARBARA
/ TITLE OF INVENTION: RECOMBINANT DNA MOLECULE COMPRISING
/ LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN 3 PHOSPHATIDYLINOSITOL
/ LINKAGE SIGNAL SEQUENCE
/ NUMBER OF SEQUENCES: 12
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/335,688
/ FILING DATE: 10-APR-1989
/ SEQ ID NO:11:
/ LENGTH: 318
5223394-11

```

```
Query Match      58.8%; Score 1363; DB 6; Length 318;
Best Local Similarity 93.4%; Pred. No. 1.9e-106;
Matches 268; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY      1 NMRGVPFRHLILVQLALIPATOGKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
      1 NMRGVPFRHLILVQLALIPATOGKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
DB      1 NMRGVPFRHLILVQLALIPATOGKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
QY      61 ILNGGSPFTKGPSTKLNDRADSRSLMDGNFPLIIKNLKIEDSTYICEVEDQKEEVQL 120
      61 ILNGGSPFTKGPSTKLNDRADSRSLMDGNFPLIIKNLKIEDSTYICEVEDQKEEVQL 120
DB      61 ILNGGSPFTKGPSTKLNDRADSRSLMDGNFPLIIKNLKIEDSTYICEVEDQKEEVQL 120
QY      121 LVFGILTANSDTHLLOQOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
      121 LVFGILTANSDTHLLOQOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
DB      121 LVFGILTANSDTHLLOQOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
QY      181 TWTCITVLONQKKEVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTSGGELMW 240
      181 TWTCITVLONQKKEVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTSGGELMW 240
DB      181 TWTCITVLONQKKEVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTSGGELMW 240
QY      241 QAEKASSSKSWITTFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOA 287
      241 QAEKASSSKSWITTFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOA 287
DB      241 QAEKASSSKSWITTFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOA 287

RESULT 36
5223394-9
; Patent No. 5223394
; APPLICANT: WALINER, BARBARA
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULE COMPRISING
; LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN 3 PHOSPHATIDYLINOSITOL
; LINKAGE SIGNAL SEQUENCE
; NUMBER OF SEQUENCES: 12
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/335,688
; FILING DATE: 10-APR-1989
; SEQ ID NO: 9
; LENGTH: 295
5223394-9

Query Match      58.6%; Score 1363; DB 6; Length 295;
Best Local Similarity 99.3%; Pred. No. 4.5e-106;
Matches 266; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 NMRGVPFRHLILVQLALIPATOGKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
      1 NMRGVPFRHLILVQLALIPATOGKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
DB      1 NMRGVPFRHLILVQLALIPATOGKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
QY      61 ILNGGSPFTKGPSTKLNDRADSRSLMDGNFPLIIKNLKIEDSTYICEVEDQKEEVQL 120
      61 ILNGGSPFTKGPSTKLNDRADSRSLMDGNFPLIIKNLKIEDSTYICEVEDQKEEVQL 120
DB      61 ILNGGSPFTKGPSTKLNDRADSRSLMDGNFPLIIKNLKIEDSTYICEVEDQKEEVQL 120
QY      121 LVFGILTANSDTHLLOQOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
      121 LVFGILTANSDTHLLOQOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
DB      121 LVFGILTANSDTHLLOQOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
QY      181 TWTCITVLONQKKEVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTSGGELMW 240
      181 TWTCITVLONQKKEVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTSGGELMW 240
DB      181 TWTCITVLONQKKEVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTSGGELMW 240
QY      241 QAEKASSSKSWITTFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOA 287
      241 QAEKASSSKSWITTFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOA 287
DB      241 QAEKASSSKSWITTFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOA 287

RESULT 37
US-08-477-4608-6
; Sequence 6, Application US/084774608
; Patent No. 6034223
; GENERAL INFORMATION:
; APPLICANT: Progenice Pharmaceuticals, Inc.
; TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
; TITLE OF INVENTION: CD4-GAMMA2 AND CD4-1GG2 IMMUNOCONJUGATES, AND USES THEREOF
```

```
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,4608
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPW/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9809
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-08-477-4608-6

Query Match      45.2%; Score 1050.5; DB 3; Length 310;
Best Local Similarity 75.4%; Pred. No. 6.8e-80;
Matches 221; Conservative 12; Mismatches 41; Indels 19; Gaps 4;

QY      1 NMRGVPFRHLILVQLALIPATOGKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
      1 NMRGVPFRHLILVQLALIPATOGKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
DB      1 NMRGVPFRHLILVQLALIPATOGKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
QY      61 ILNGGSPFTKGPSTKLNDRADSRSLMDGNFPLIIKNLKIEDSTYICEVEDQKEEVQL 120
      61 ILNGGSPFTKGPSTKLNDRADSRSLMDGNFPLIIKNLKIEDSTYICEVEDQKEEVQL 120
DB      61 ILNGGSPFTKGPSTKLNDRADSRSLMDGNFPLIIKNLKIEDSTYICEVEDQKEEVQL 120
QY      121 LVFGILTANSDTHLLOQOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
      121 LVFGILTANSDTHLLOQOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
DB      121 LVFGILTANSDTHLLOQOSLTLTLESPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSG 180
QY      181 TWTCITVLONQKKEVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTSGGELMW 240
      181 TWTCITVLONQKKEVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTSGGELMW 240
DB      181 TWTCITVLONQKKEVEFKIDIVLAFOKASSIVYKKEGEQVEFSFPLAFVTEKLTSGGELMW 240
QY      241 QAEKASSSKSWITTFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOA 287
      241 QAEKASSSKSWITTFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOA 287
DB      241 QAEKASSSKSWITTFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOA 287

RESULT 38
US-08-379-516-6
; Sequence 6, Application US/08379516
; Patent No. 6083478
; GENERAL INFORMATION:
; APPLICANT: Allaway, Graham P.
; TITLE OF INVENTION: Moieity-Conjugated CD4-Gamma2 and CD4-1GG2
```

TITLE OF INVENTION: Immunocjugates and Uses thereof
 FILE REFERENCE: 41215-A-PCT-US
 CURRENT APPLICATION NUMBER: US/08/379,516
 CURRENT FILING DATE: 1996-06-10
 EARLIER APPLICATION NUMBER: PCT/US93/07422
 EARLIER FILING DATE: 1993-08-06
 EARLIER APPLICATION NUMBER: 07/927,931
 EARLIER FILING DATE: 1992-08-07
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO: 6
 LENGTH: 310
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-08-379-516-6

Query Match 45.2%; Score 1050.5; DB 3; Length 310;
 Best Local Similarity 75.4%; Pred. No. 6.8e-80;
 Matches 221; Conservative 12; Mismatches 41; Indels 19; Gaps 4;

Qy 1 MNRGVPFRHLLVQLALLPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHMKNNOIK 60
 Db 1 MNRGVPFRHLLVQLALLPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHMKNNOIK 60
 Qy 61 ILNGGSFLLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVQL 120
 Db 61 ILNGGSFLLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVQL 120
 Qy 121 LVFGLTANSPTHLQGSLLTLLESPPGSSPVQCSRGRKNIQGGKTLVSQLELDQSG 180
 Db 121 LVFGLTANSPTHLQGSLLTLLESPPGSSPVQCSRGRKNIQGGKTLVSQLELDQSG 180
 Qy 181 TWTCVQLNQKQKVEFKIDIVLAFOKA--SSIVYKKEGEQVEFPLAFVETKLTG---- 234
 Db 181 TWTCVQLNQKQKVEFKIDIVLAFOKA--SSIVYKKEGEQVEFPLAFVETKLTG---- 234
 Qy 235 SGEIWMQAEASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTPQA 287
 Db 235 SGEIWMQAEASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTPQA 287
 Qy 239 EAKVQMKVDNALQSGN-----SQESVTEQDSKDSSTYLSSTLTLSKA 280
 Db 239 EAKVQMKVDNALQSGN-----SQESVTEQDSKDSSTYLSSTLTLSKA 280

RESULT 39
 US-09-329-916-6
 Sequence 6, Application US/09329916
 Patent No. 6177549
 GENERAL INFORMATION:
 APPLICANT: Progenics Pharmaceuticals, Inc.
 TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
 TITLE OF INVENTION: CD4-GAMMA2 AND CD4-IgG2 IMMUNOCONJUGATES, AND USES THEREOF
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham
 STREET: 30 Rockefeller Plaza
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10112
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/329,916
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/477,460
 FILING DATE: 07-JUN-1995
 APPLICATION NUMBER: US 07/927,931
 FILING DATE: 07-AUG-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.

REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPW/AJM
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 977-9550
 TELEFAX: (212) 977-9809
 TELEX: 422523 COOP UI
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 310 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: homo sapien
 CELL TYPE: lymphocyte
 US-09-329-916-6

Query Match 45.2%; Score 1050.5; DB 3; Length 310;
 Best Local Similarity 75.4%; Pred. No. 6.8e-80;
 Matches 221; Conservative 12; Mismatches 41; Indels 19; Gaps 4;

Qy 1 MNRGVPFRHLLVQLALLPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHMKNNOIK 60
 Db 1 MNRGVPFRHLLVQLALLPAATQGNKVVLGKGGDTVELTCTASOKKSIOFHMKNNOIK 60
 Qy 61 ILNGGSFLLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVQL 120
 Db 61 ILNGGSFLLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVQL 120
 Qy 121 LVFGLTANSPTHLQGSLLTLLESPPGSSPVQCSRGRKNIQGGKTLVSQLELDQSG 180
 Db 121 LVFGLTANSPTHLQGSLLTLLESPPGSSPVQCSRGRKNIQGGKTLVSQLELDQSG 180
 Qy 181 TWTCVQLNQKQKVEFKIDIVLAFOKA--SSIVYKKEGEQVEFPLAFVETKLTG---- 234
 Db 181 TWTCVQLNQKQKVEFKIDIVLAFOKA--SSIVYKKEGEQVEFPLAFVETKLTG---- 234
 Qy 235 SGEIWMQAEASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTPQA 287
 Db 235 SGEIWMQAEASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTPQA 287
 Qy 239 EAKVQMKVDNALQSGN-----SQESVTEQDSKDSSTYLSSTLTLSKA 280
 Db 239 EAKVQMKVDNALQSGN-----SQESVTEQDSKDSSTYLSSTLTLSKA 280

RESULT 40
 US-08-485-372A-6
 Sequence 6, Application US/08485372A
 Patent No. 6187748
 GENERAL INFORMATION:
 APPLICANT: Beaudry, Gary A.
 TITLE OF INVENTION: CD4-GAMMA2 CD4-IgG2 CHIMERAS
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham LLP
 STREET: 1185 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/485,372A
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/476,227
 FILING DATE: 07-JUN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.

REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 37690-II-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-08-485-372A-6

Query Match 45.2%; Score 1050.5; DB 3; Length 310;
Best Local Similarity 75.4%; Pred. No. 6.8e-80;
Matches 221; Conservative 12; Mismatches 41; Indels 19; Gaps 4;

QY 1 MNRGVPRHLLLVQLALPPATQGNKVVLGKKGDTVELTCTASQKSIQFHMKNNOIK 60
DB 1 MNRGVPRHLLLVQLALPPATQGNKVVLGKKGDTVELTCTASQKSIQFHMKNNOIK 60
QY 61 ILGNQSFLLTGKPSKLNDRADSRSLMDQGNPPLIKNLKIEDSTYICEVEDQKEEYOL 120
DB 61 ILGNQSFLLTGKPSKLNDRADSRSLMDQGNPPLIKNLKIEDSTYICEVEDQKEEYOL 120
QY 121 LVFGLTANSDTHLQOGSLTTLTSPGSSPVOCRSRPGKNIQGGKTLVSQLELODSG 180
DB 121 LVFGLTANSDTHLQOGSLTTLTSPGSSPVOCRSRPGKNIQGGKTLVSQLELODSG 180
QY 181 TWTCVTLQNKQKVEFKIDIVLAFOKA--SSIVYKKEGEQVEFSPLAFTVEKLTG--- 234
DB 181 TWTCVTLQNKQKVEFKIDIVLAFTVAAPSVFIFFPSDEQLKSG--TASVCLLNPFYR 238
QY 235 SGEIWMQERASSKSWITFDLKNKEVSKRYTQDPKLOMGKKLPLHLTPQA 287
DB 235 SGEIWMQERASSKSWITFDLKNKEVSKRYTQDPKLOMGKKLPLHLTPQA 287

RESULT 41
US-09-409-006A-6
Sequence 6, Application US/09409006A
Patent No. 6342586
GENERAL INFORMATION:
APPLICANT: Progenice Pharmaceuticals, Inc.
TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/409,006A
FILING DATE: 29-SEP-1999
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.

REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPM/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9809
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-09-409-006A-6

Query Match 45.2%; Score 1050.5; DB 4; Length 310;
Best Local Similarity 75.4%; Pred. No. 6.8e-80;
Matches 221; Conservative 12; Mismatches 41; Indels 19; Gaps 4;

QY 1 MNRGVPRHLLLVQLALPPATQGNKVVLGKKGDTVELTCTASQKSIQFHMKNNOIK 60
DB 1 MNRGVPRHLLLVQLALPPATQGNKVVLGKKGDTVELTCTASQKSIQFHMKNNOIK 60
QY 61 ILGNQSFLLTGKPSKLNDRADSRSLMDQGNPPLIKNLKIEDSTYICEVEDQKEEYOL 120
DB 61 ILGNQSFLLTGKPSKLNDRADSRSLMDQGNPPLIKNLKIEDSTYICEVEDQKEEYOL 120
QY 121 LVFGLTANSDTHLQOGSLTTLTSPGSSPVOCRSRPGKNIQGGKTLVSQLELODSG 180
DB 121 LVFGLTANSDTHLQOGSLTTLTSPGSSPVOCRSRPGKNIQGGKTLVSQLELODSG 180
QY 181 TWTCVTLQNKQKVEFKIDIVLAFOKA--SSIVYKKEGEQVEFSPLAFTVEKLTG--- 234
DB 181 TWTCVTLQNKQKVEFKIDIVLAFTVAAPSVFIFFPSDEQLKSG--TASVCLLNPFYR 238
QY 235 SGEIWMQERASSKSWITFDLKNKEVSKRYTQDPKLOMGKKLPLHLTPQA 287
DB 235 SGEIWMQERASSKSWITFDLKNKEVSKRYTQDPKLOMGKKLPLHLTPQA 287

RESULT 42
US-08-484-681-6
Sequence 6, Application US/08484681
Patent No. 6451313
GENERAL INFORMATION:
APPLICANT: Beaudry, Gary A.
TITLE OF INVENTION: CD4-GAMMA2 CD4-IgG2 CHIMERAS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,681
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 37690-II-B
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-08-484-681-6

Query Match 45.2%; Score 1050.5; DB 4; Length 310;
Best Local Similarity 75.4%; Pred. No. 6.8e-80;
Matches 221; Conservative 12; Mismatches 41; Indels 19; Gaps 4;

QY 1 MNRGVPFRHLVLVQLALPPATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNOIK 60
DB 1 MNRGVPFRHLVLVQLALPPATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNOIK 60
QY 61 ILGNQGSFLTKGSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180
DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180
QY 181 TWTCVQLNQKKEFKIDIVLAFQKA--SSIVYKKEGQVEFSPLAFTVEKLTG---- 234
DB 181 TWTCVQLNQKKEFKIDIVLAFQKA--SSIVYKKEGQVEFSPLAFTVEKLTG---- 234
QY 235 SGEIWMQAEARASSKSMITFDLKNKEVSVKRYTQDPLQMGKPLHLITLPOA 287
DB 235 SGEIWMQAEARASSKSMITFDLKNKEVSVKRYTQDPLQMGKPLHLITLPOA 287
QY 239 EAKVQWKVDNALQSGN-----SQESVTEQDSKDSYSLSTLTLSKA 280
DB 239 EAKVQWKVDNALQSGN-----SQESVTEQDSKDSYSLSTLTLSKA 280

RESULT 43
PCT-US93-07422-6
Sequence 6, Application PC/TUS9307422

GENERAL INFORMATION:
APPLICANT: Progenics Pharmaceuticals, Inc.
TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
TITLE OF INVENTION: CD4-GAMMA2 AND CD4-IGG2 IMMUNOCONJUGATES, AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07422
FILING DATE: 19930806
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPM/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550

TELEFAX: (212) 977-9809
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
PCT-US93-07422-6

Query Match 45.2%; Score 1050.5; DB 5; Length 310;
Best Local Similarity 75.4%; Pred. No. 6.8e-80;
Matches 221; Conservative 12; Mismatches 41; Indels 19; Gaps 4;

QY 1 MNRGVPFRHLVLVQLALPPATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNOIK 60
DB 1 MNRGVPFRHLVLVQLALPPATQGNKVVLGKGGDTVELTCTASOKKSIQFHMKNNOIK 60
QY 61 ILGNQGSFLTKGSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180
DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180
QY 181 TWTCVQLNQKKEFKIDIVLAFQKA--SSIVYKKEGQVEFSPLAFTVEKLTG---- 234
DB 181 TWTCVQLNQKKEFKIDIVLAFQKA--SSIVYKKEGQVEFSPLAFTVEKLTG---- 234
QY 235 SGEIWMQAEARASSKSMITFDLKNKEVSVKRYTQDPLQMGKPLHLITLPOA 287
DB 235 SGEIWMQAEARASSKSMITFDLKNKEVSVKRYTQDPLQMGKPLHLITLPOA 287
QY 239 EAKVQWKVDNALQSGN-----SQESVTEQDSKDSYSLSTLTLSKA 280
DB 239 EAKVQWKVDNALQSGN-----SQESVTEQDSKDSYSLSTLTLSKA 280

RESULT 44

US-08-477-460B-4
Sequence 4, Application US/08477460B
Patent No. 6034223

GENERAL INFORMATION:
APPLICANT: Progenics Pharmaceuticals, Inc.
TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
TITLE OF INVENTION: CD4-GAMMA2 AND CD4-IGG2 IMMUNOCONJUGATES, AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,460B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPM/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550

TELEFAX: (212) 977-9809
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-08-477-460B-4

Query Match 45.0%; Score 1046; DB 3; Length 530;
Best Local Similarity 93.4%; Pred. No. 3.6e-79;
Matches 21; Conservative 2; Mismatches 3; Indels 10; Gaps 3;

QY 1 NMRGVPFRHLILVQLALPPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNSNOIK 60
DB 1 NMRGVPFRHLILVQLALPPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNSNOIK 60
QY 61 ILNGGSLFTGKPSKLNDRADSRSLMDQGNFPLIIKULKIEDSDTYICEVEDQKEVQL 120
DB 61 ILNGGSLFTGKPSKLNDRADSRSLMDQGNFPLIIKULKIEDSDTYICEVEDQKEVQL 120
QY 121 LVFGILTANSDTHLLOQGSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180
DB 121 LVFGILTANSDTHLLOQGSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180
QY 181 TWTCTVLONOKKVEFKIDIVLAFQKASIVYKKEGEVFEFSPFLA 226
DB 181 TWTCTVLONOKKVEFKIDIVLAF--AST-----KGPSTV---FPLA 216

RESULT 45
US-08-379-516-4
Sequence 4, Application US/08379516
Patent No. 6083478
GENERAL INFORMATION:
APPLICANT: Allaway, Graham P.
TITLE OF INVENTION: No. 6083478-peptidyl moiety-conjugated CD4-Gamma2 and CD4-1G62
TITLE OF INVENTION: Immunocjugates and Uses Thereof
FILE REFERENCE: 41215-A-PCT-US
CURRENT APPLICATION NUMBER: US/08/379,516
CURRENT FILING DATE: 1996-06-10
EARLIER APPLICATION NUMBER: PCT/US93/07422
EARLIER FILING DATE: 1993-08-06
EARLIER APPLICATION NUMBER: 07/927,931
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 4
LENGTH: 530
TYPE: PRT
ORGANISM: Homo sapiens
US-08-379-516-4

Query Match 45.0%; Score 1046; DB 3; Length 530;
Best Local Similarity 93.4%; Pred. No. 3.6e-79;
Matches 21; Conservative 2; Mismatches 3; Indels 10; Gaps 3;

QY 1 NMRGVPFRHLILVQLALPPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNSNOIK 60
DB 1 NMRGVPFRHLILVQLALPPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNSNOIK 60
QY 61 ILNGGSLFTGKPSKLNDRADSRSLMDQGNFPLIIKULKIEDSDTYICEVEDQKEVQL 120
DB 61 ILNGGSLFTGKPSKLNDRADSRSLMDQGNFPLIIKULKIEDSDTYICEVEDQKEVQL 120
QY 121 LVFGILTANSDTHLLOQGSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180
DB 121 LVFGILTANSDTHLLOQGSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180

DB 121 LVFGILTANSDTHLLOQGSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180
QY 181 TWTCTVLONOKKVEFKIDIVLAFQKASIVYKKEGEVFEFSPFLA 226
DB 181 TWTCTVLONOKKVEFKIDIVLAF--AST-----KGPSTV---FPLA 216

RESULT 46
US-09-329-916-4
Sequence 4, Application US/09329916
Patent No. 6177549
GENERAL INFORMATION:
APPLICANT: Progenics Pharmaceuticals, Inc.
TITLE OF INVENTION: CD4-GAMMA2 AND CD4-1G62 IMMUNOCJUGATES, AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/329,916
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/477,460
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPM/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9809
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-09-329-916-4

Query Match 45.0%; Score 1046; DB 3; Length 530;
Best Local Similarity 93.4%; Pred. No. 3.6e-79;
Matches 21; Conservative 2; Mismatches 3; Indels 10; Gaps 3;

QY 1 NMRGVPFRHLILVQLALPPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNSNOIK 60
DB 1 NMRGVPFRHLILVQLALPPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNSNOIK 60
QY 61 ILNGGSLFTGKPSKLNDRADSRSLMDQGNFPLIIKULKIEDSDTYICEVEDQKEVQL 120
DB 61 ILNGGSLFTGKPSKLNDRADSRSLMDQGNFPLIIKULKIEDSDTYICEVEDQKEVQL 120
QY 121 LVFGILTANSDTHLLOQGSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180
DB 121 LVFGILTANSDTHLLOQGSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180

Qy 181 TWCTCTVLQNKQKVEFKIDIVLAFQKASSIYKKEGEVSFPPLA 226
Db 181 TWCTCTVLQNKQKVEFKIDIVLAF--AST-----KGPSV---FPLA 216

RESULT 47
US-08-485-372A-4
Sequence 4, Application US/08485372A
Patent No. 6187748
GENERAL INFORMATION:
APPLICANT: Beaudry, Gary A.
APPLICANT: Maddon, Paul J.
TITLE OF INVENTION: CD4-GAMMA2 CD4-IGG2 CHIMERAS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,372A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,227
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 37690-II-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-08-485-372A-4

Query Match 45.0%; Score 1046; DB 3; Length 530;
Best Local Similarity 93.4%; Pred. No. 3,6e-79;
Matches 211; Conservative 2; Mismatches 3; Indels 10; Gaps 3;

Qy 1 MNRGVPRHLLVQLALLPAATQGNKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
Db 1 MNRGVPRHLLVQLALLPAATQGNKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
Qy 61 ILGNQGSFLLTGKPSKLNDRADSRSLWDQGNFPLIINKLKIESDPTVCEVEDQKEEVOL 120
Db 61 ILGNQGSFLLTGKPSKLNDRADSRSLWDQGNFPLIINKLKIESDPTVCEVEDQKEEVOL 120
Qy 121 LVFGLTANSDTHLLQGSQSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLSSVQLDSDG 180
Db 121 LVFGLTANSDTHLLQGSQSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLSSVQLDSDG 180
Qy 181 TWCTCTVLQNKQKVEFKIDIVLAFQKASSIYKKEGEVSFPPLA 226
Db 181 TWCTCTVLQNKQKVEFKIDIVLAF--AST-----KGPSV---FPLA 216

RESULT 48
US-09-409-006A-4
Sequence 4, Application US/09409006A
Patent No. 6342586
GENERAL INFORMATION:
APPLICANT: Progenics Pharmaceuticals, Inc.
TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/409,006A
FILING DATE: 29-SEP-1999
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPW/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9809
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-09-409-006A-4

Query Match 45.0%; Score 1046; DB 4; Length 530;
Best Local Similarity 93.4%; Pred. No. 3,6e-79;
Matches 211; Conservative 2; Mismatches 3; Indels 10; Gaps 3;

Qy 1 MNRGVPRHLLVQLALLPAATQGNKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
Db 1 MNRGVPRHLLVQLALLPAATQGNKVVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60
Qy 61 ILGNQGSFLLTGKPSKLNDRADSRSLWDQGNFPLIINKLKIESDPTVCEVEDQKEEVOL 120
Db 61 ILGNQGSFLLTGKPSKLNDRADSRSLWDQGNFPLIINKLKIESDPTVCEVEDQKEEVOL 120
Qy 121 LVFGLTANSDTHLLQGSQSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLSSVQLDSDG 180
Db 121 LVFGLTANSDTHLLQGSQSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLSSVQLDSDG 180
Qy 181 TWCTCTVLQNKQKVEFKIDIVLAFQKASSIYKKEGEVSFPPLA 226
Db 181 TWCTCTVLQNKQKVEFKIDIVLAF--AST-----KGPSV---FPLA 216

RESULT 49
US-08-484-681-4
Sequence 4, Application US/08484681

```

; Patent No. 6451313
; GENERAL INFORMATION:
; APPLICANT: Beaudry, Gary A.
; APPLICANT: Maddon, Paul J.
; TITLE OF INVENTION: CD4-GAMMA2 CD4-IgG2 CHIMERAS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,681
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 37690-II-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: homo sapien
; CELL TYPE: lymphocyte
; US-08-484-681-4

Query Match          45.0%; Score 1046; DB 4; Length 530;
Best Local Similarity 93.4%; Pred. No. 3,66-79;
Matches 21; Conservative 2; Mismatches 3; Indels 10; Gaps 3;

QY 1 MNRGVPRFRLHLVQLALPPATOGNKVVLGKKGDVVELTCTASQKSIQPHMKNNOIK 60
DB 1 MNRGVPRFRLHLVQLALPPATOGNKVVLGKKGDVVELTCTASQKSIQPHMKNNOIK 60
QY 61 ILNGGSFLTKGPKSLNDRADSRSLMDQGNFPLIIKIKLIEDSDTYICEVEDQKEEYQL 120
DB 61 ILNGGSFLTKGPKSLNDRADSRSLMDQGNFPLIIKIKLIEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSDPHLLQOGSLTTLTSSPPSSPVOCRSRPGKNIOGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDPHLLQOGSLTTLTSSPPSSPVOCRSRPGKNIOGGKTLVSQLELDQSG 180
QY 181 TWTCTVLONQKKEFKIDIVLAFQKASSIVYKKEGEQVEFSFPLA 226
DB 181 TWTCTVLONQKKEFKIDIVLAF--AST-----KGPSV---FPLA 216

RESULT 50
PCT-US93-07422-4
; Sequence 4, Application PC/TUS9307422
; GENERAL INFORMATION:
; APPLICANT: Progenics Pharmaceuticals, Inc.
; TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA

```

```

; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07422
; FILING DATE: 19930806
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,931
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPM/ADM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: 422523 COOP UI
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: homo sapien
; CELL TYPE: lymphocyte
; PCT-US93-07422-4

Query Match          45.0%; Score 1046; DB 5; Length 530;
Best Local Similarity 93.4%; Pred. No. 3,66-79;
Matches 21; Conservative 2; Mismatches 3; Indels 10; Gaps 3;

QY 1 MNRGVPRFRLHLVQLALPPATOGNKVVLGKKGDVVELTCTASQKSIQPHMKNNOIK 60
DB 1 MNRGVPRFRLHLVQLALPPATOGNKVVLGKKGDVVELTCTASQKSIQPHMKNNOIK 60
QY 61 ILNGGSFLTKGPKSLNDRADSRSLMDQGNFPLIIKIKLIEDSDTYICEVEDQKEEYQL 120
DB 61 ILNGGSFLTKGPKSLNDRADSRSLMDQGNFPLIIKIKLIEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSDPHLLQOGSLTTLTSSPPSSPVOCRSRPGKNIOGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDPHLLQOGSLTTLTSSPPSSPVOCRSRPGKNIOGGKTLVSQLELDQSG 180
QY 181 TWTCTVLONQKKEFKIDIVLAFQKASSIVYKKEGEQVEFSFPLA 226
DB 181 TWTCTVLONQKKEFKIDIVLAF--AST-----KGPSV---FPLA 216

RESULT 51
US-08-477-460B-2
; Sequence 2, Application US/08477460B
; Patent No. 6034223
; GENERAL INFORMATION:
; APPLICANT: Progenics Pharmaceuticals, Inc.
; TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA

```

```
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,460B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPW/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9809
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-08-477-460B-2
```

```
Query Match 44.9%; Score 1045; DB 3; Length 432;
Best Local Similarity 98.5%; Pred. No.3.2e-79;
Matches 203; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 MNRGVPFRHLVLVQLALPPATQGNKVVLGKGDVVELTCTASQKSIQPHMKNNSQIK 60
DB 1 MNRGVPFRHLVLVQLALPPATQGNKVVLGKGDVVELTCTASQKSIQPHMKNNSQIK 60
QY 61 ILNGQSFLLTGKPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGQSFLLTGKPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTNSDTHLLQGOSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTNSDTHLLQGOSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTVLQNGKVEFKIDIVLAFQK 206
DB 181 TWCTVLQNGKVEFKIDIVLAFER 206

RESULT 52
US-08-379-516-2
Sequence 2, Application US/08379516
Patent No. 6083478
GENERAL INFORMATION:
APPLICANT: Allaway, Graham P.
APPLICANT: Madden, Paul J.
TITLE OF INVENTION: No. 6083478-Peptide1 Moieity-Conjugated CD4-Gamma2 and CD4-IgG2
FILE REFERENCE: 41215-A-PCT-US
CURRENT APPLICATION NUMBER: US/08/379,516
CURRENT FILING DATE: 1996-06-10
EARLIER APPLICATION NUMBER: PCT/US93/07422
EARLIER FILING DATE: 1993-08-06
EARLIER APPLICATION NUMBER: 07/927,931
EARLIER FILING DATE: 1992-08-07
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
```

```
LENGTH: 432
TYPE: PRT
ORGANISM: Homo sapiens
US-08-379-516-2
```

```
Query Match 44.9%; Score 1045; DB 3; Length 432;
Best Local Similarity 98.5%; Pred. No.3.2e-79;
Matches 203; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 MNRGVPFRHLVLVQLALPPATQGNKVVLGKGDVVELTCTASQKSIQPHMKNNSQIK 60
DB 1 MNRGVPFRHLVLVQLALPPATQGNKVVLGKGDVVELTCTASQKSIQPHMKNNSQIK 60
QY 61 ILNGQSFLLTGKPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGQSFLLTGKPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTNSDTHLLQGOSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTNSDTHLLQGOSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWCTVLQNGKVEFKIDIVLAFQK 206
DB 181 TWCTVLQNGKVEFKIDIVLAFER 206
```

```
RESULT 53
US-09-329-916-2
Sequence 2, Application US/09329916
Patent No. 6177549
GENERAL INFORMATION:
APPLICANT: Progenics Pharmaceuticals, Inc.
TITLE OF INVENTION: NON-PEPTIDYL MOIEITY-CONJUGATED
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/329,916
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/477,460
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPW/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9809
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: homo sapien
```

CELL TYPE: lymphocyte
US-09-329-916-2

Query Match 44.9%; Score 1045; DB 3; Length 432;

Best Local Similarity 98.5%; Pred. No. 3.2e-79;
Matches 203; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKSIQPFHMKNSNOIK 60
DB 1 MNRGVPFRHLLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKSIQPFHMKNSNOIK 60
QY 61 ILGNQGSFLTKGSPSKLNRADSRSLMDQGNPPLIIKNLKIETDPTIICEVEDQKEEYQL 120
DB 61 ILGNQGSFLTKGSPSKLNRADSRSLMDQGNPPLIIKNLKIETDPTIICEVEDQKEEYQL 120
QY 121 LVFGILTANSDTHLQGSQSLTTLSPSGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGILTANSDTHLQGSQSLTTLSPSGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNKVKVEFKIDIVLAFOK 206
DB 181 TWTCTVLQNKVKVEFKIDIVLAFOK 206

RESULT 54

US-08-485-372A-2
Sequence 2, Application US/08485372A

Patent No. 6187748

GENERAL INFORMATION:

APPLICANT: Beaudry, Gary A.

APPLICANT: Maddon, Paul J.

TITLE OF INVENTION: CD4-GAMMA2 CD4-IGG2 CHIMERAS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/485.372A

CLASSIFICATION: 435

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,227

FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 37690-II-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 432 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: protein

ORGANISM: homo sapien

CELL TYPE: lymphocyte

US-08-485-372A-2

Query Match 44.9%; Score 1045; DB 3; Length 432;

Best Local Similarity 98.5%; Pred. No. 3.2e-79;
Matches 203; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKSIQPFHMKNSNOIK 60
DB 1 MNRGVPFRHLLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKSIQPFHMKNSNOIK 60
QY 61 ILGNQGSFLTKGSPSKLNRADSRSLMDQGNPPLIIKNLKIETDPTIICEVEDQKEEYQL 120
DB 61 ILGNQGSFLTKGSPSKLNRADSRSLMDQGNPPLIIKNLKIETDPTIICEVEDQKEEYQL 120
QY 121 LVFGILTANSDTHLQGSQSLTTLSPSGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGILTANSDTHLQGSQSLTTLSPSGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNKVKVEFKIDIVLAFOK 206
DB 181 TWTCTVLQNKVKVEFKIDIVLAFOK 206

RESULT 55

US-09-409-006A-2
Sequence 2, Application US/09409006A

Patent No. 6342586

GENERAL INFORMATION:

APPLICANT: Progenics Pharmaceuticals, Inc.

TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/409.006A

FILING DATE: 29-SEP-1999

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/927,931

FILING DATE: 07-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPW/AJM

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 977-9550

TELEFAX: (212) 977-9809

TELEX: 422523 COOP UT

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 432 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: protein

ORGANISM: homo sapien

CELL TYPE: lymphocyte

US-09-409-006A-2

Query Match 44.9%; Score 1045; DB 4; Length 432;
Best Local Similarity 98.5%; Pred. No. 3.2e-79;
Matches 203; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKSIQPFHMKNSNOIK 60

Db 1 MNRGVPFRHLLVLTQALLPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNSNOIK 60
Qy 61 ILNGQGSFLLTGKPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEVQL 120
Db 61 ILNGQGSFLLTGKPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEVQL 120
Qy 121 LVFGLTANSDPHTLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Db 121 LVFGLTANSDPHTLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Qy 181 TWCTVVLQONOKKVEFKIDIVLAFQK 206
Db 181 TWCTVVLQONOKKVEFKIDIVLAFER 206

RESULT 56
US-08-484-681-2
; Sequence 2, Application US/08484681
; Patent No. 6451313
; GENERAL INFORMATION:
; APPLICANT: Beaudry, Gary A.
; APPLICANT: Maddon, Paul J.
; TITLE OF INVENTION: CD4-GAMMA2 CD4-19G2 CHIMERAS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,681
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 37690-II-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: homo sapien
; CELL TYPE: lymphocyte
; US-08-484-681-2

Query Match 44.9%; Score 1045; DB 4; Length 432;
Best Local Similarity 98.5%; Pred. No. 3.2e-79;
Matches 203; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MNRGVPFRHLLVLTQALLPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNSNOIK 60
Db 1 MNRGVPFRHLLVLTQALLPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNSNOIK 60
Qy 61 ILNGQGSFLLTGKPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEVQL 120
Db 61 ILNGQGSFLLTGKPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEVQL 120
Qy 121 LVFGLTANSDPHTLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Db 121 LVFGLTANSDPHTLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Qy 181 TWCTVVLQONOKKVEFKIDIVLAFQK 206
Db 181 TWCTVVLQONOKKVEFKIDIVLAFER 206

Qy 121 LVFGLTANSDPHTLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Db 121 LVFGLTANSDPHTLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Qy 181 TWCTVVLQONOKKVEFKIDIVLAFQK 206
Db 181 TWCTVVLQONOKKVEFKIDIVLAFER 206

RESULT 57
PCT-US93-07422-2
; Sequence 2, Application PC/TUS9307422
; GENERAL INFORMATION:
; APPLICANT: Progenics Pharmaceuticals, Inc.
; TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
; TITLE OF INVENTION: CD4-GAMMA2 AND CD4-19G2 IMMUNOCONJUGATES, AND USES THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07422
; FILING DATE: 19930806
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,931
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPW/AJM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 977-9809
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: homo sapien
; CELL TYPE: lymphocyte
; PCT-US93-07422-2

Query Match 44.9%; Score 1045; DB 5; Length 432;
Best Local Similarity 98.5%; Pred. No. 3.2e-79;
Matches 203; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MNRGVPFRHLLVLTQALLPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNSNOIK 60
Db 1 MNRGVPFRHLLVLTQALLPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNSNOIK 60
Qy 61 ILNGQGSFLLTGKPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEVQL 120
Db 61 ILNGQGSFLLTGKPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEVQL 120
Qy 121 LVFGLTANSDPHTLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Db 121 LVFGLTANSDPHTLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Qy 181 TWCTVVLQONOKKVEFKIDIVLAFQK 206
Db 181 TWCTVVLQONOKKVEFKIDIVLAFER 206

Db 181 TWTCTVLGNQKKVEFKIDIVLA 206

```
|||||
RESULT 58
US-08-284-391B-31
/ Sequence 31, Application US/08284391B
/ Patent No. 5851828
/ GENERAL INFORMATION:
/ APPLICANT: Seed, Brian
/ APPLICANT: Banapour, Babak
/ APPLICANT: Romeo, Charles
/ APPLICANT: Kolanus, Waldemar
/ TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
/ TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
/ NUMBER OF SEQUENCES: 53
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Clark & Ebling LLP
/ STREET: 176 Federal Street
/ CITY: Boston
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02110
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FASTSEQ for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/284,391B
/ FILING DATE: 02-AUG-1994
/ CLASSIFICATION: 514
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/195,395
/ FILING DATE: 14-FEB-1994
/ APPLICATION NUMBER: 07/847,566
/ FILING DATE: 06-MAR-1992
/ APPLICATION NUMBER: 07/665,961
/ FILING DATE: 07-MAR-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ebling, Karen L.
/ REGISTRATION NUMBER: 35,238
/ REFERENCE/DOCKET NUMBER: 00786/247001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617-428-0200
/ TELEFAX: 617-428-7045
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 31:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 203 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-284-391B-31

Query Match 44.8%; Score 1041; DB 2; Length 203;
Best Local Similarity 100.0%; Pred. No. 2.3e-79;
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLVQLALPAPATQGNKVVLGKKGDTVELTCTASQKSIQPHMKNNSQIK 60
DB 1 MNRGVPFRHLVLVQLALPAPATQGNKVVLGKKGDTVELTCTASQKSIQPHMKNNSQIK 60
QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEEYQL 120
DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGILTANSDTHLLQGQSLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGILTANSDTHLLQGQSLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 121 LVFGILTANSDTHLLQGQSLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGILTANSDTHLLQGQSLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCTVLGNQKKVEFKIDIVLA 203
```

Db 181 TWTCTVLGNQKKVEFKIDIVLA 203

```
|||||
RESULT 59
US-09-218-950-31
/ Sequence 31, Application US/09218950
/ Patent No. 6284240
/ GENERAL INFORMATION:
/ APPLICANT: Seed, Brian
/ APPLICANT: Banapour, Babak
/ APPLICANT: Romeo, Charles
/ APPLICANT: Kolanus, Waldemar
/ TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
/ TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
/ NUMBER OF SEQUENCES: 53
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Clark & Ebling LLP
/ STREET: 176 Federal Street
/ CITY: Boston
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02110
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FASTSEQ for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/218,950
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/284,391
/ FILING DATE: 02-AUG-1994
/ APPLICATION NUMBER: 08/195,395
/ FILING DATE: 14-FEB-1994
/ APPLICATION NUMBER: 07/847,566
/ FILING DATE: 06-MAR-1992
/ APPLICATION NUMBER: 07/665,961
/ FILING DATE: 07-MAR-1991
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ebling, Karen L.
/ REGISTRATION NUMBER: 35,238
/ REFERENCE/DOCKET NUMBER: 00786/247001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617-428-0200
/ TELEFAX: 617-428-7045
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 31:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 203 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-09-218-950-31

Query Match 44.8%; Score 1041; DB 3; Length 203;
Best Local Similarity 100.0%; Pred. No. 2.3e-79;
Matches 203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLVQLALPAPATQGNKVVLGKKGDTVELTCTASQKSIQPHMKNNSQIK 60
DB 1 MNRGVPFRHLVLVQLALPAPATQGNKVVLGKKGDTVELTCTASQKSIQPHMKNNSQIK 60
QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEEYQL 120
DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGILTANSDTHLLQGQSLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGILTANSDTHLLQGQSLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 121 LVFGILTANSDTHLLQGQSLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGILTANSDTHLLQGQSLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
```


Qy	119	OLIVFGLTASDPHLLQGGSLTLLTLES-PCGSSSVQCRBPRKGNIGQKGLTSVQLEQ	177
Db	121	ELWFRATYFNGTLLQGGSLTLLTLDNPKVSDPPIECRKKSSINVDVAFSTHSLRIQ	180
Qy	178	DSGTMCTVTLOQOKKVEFKIDIVLAQKASIIYKKEGQVEFSEPLAFVETKLTGSGE	237
Db	181	DSGIMNCTVTLLNQKSHSFDKMLSVLQPSAGNSITAYKSEGSABFSPFLNIGSESL--QGE	238
Qy	238	LMWAEEAASSSKSIITFDLKKKEYSVKRVYQDPFLQNGKKLPLHLTLPQALPOYASSGN	297
Db	239	LRWAEEAPSSQSQITSLSKQKQKSVKSSNPFOJSETLPLTLIDIPQVSLQFASQGN	298
Qy	298	TLALEAKTGKLRQEVNLVVMRATQLOKN-LTCEWGPSTPKMLSLKLEWKAAYSKREK	356
Db	299	TLTLD--RGILYQEVNLVVMKVPQDPBSNTLTCEWGPSTPKMLILKQENQEARVSRQEK	356
Qy	357	PWVYVLANPBAQMOCLLSDSGQVLLIESNIKYL	387
Db	357	VIQVQAEPAQWOCLLSEGEVVKMDSIKIQL	387

RESULT 62
US-08-394-442B-8
; Sequence 8, Application US/08394442B

```

1  GENERAL INFORMATION:
2  APPLICANT: Hercend, Thierry
3  APPLICANT: Triebel, Frederic
4  TITLE OF INVENTION: New Proteins Produced By Human
5  TITLE OF INVENTION: Lymphocytes, DNA Sequences Encoding These Proteins And
6  TITLE OF INVENTION: Their Pharmaceutical And Biological Uses
7  NUMBER OF SEQUENCES: 11
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
10 STREET: 419 Seventh Street, N.W., Suite 400
11 CITY: Washington
12 STATE: D.C.
13 COUNTRY: U.S.A.
14 ZIP: 20004
15
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: Patent In Release #1.0, Version #1.25
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/08/394,442B
23 FILING DATE: 24-FEB-1995
24 CLASSIFICATION: 424
25
26 ATTORNEY/AGENT INFORMATION:
27 NAME: BROWDY, Roger L.
28 REGISTRATION NUMBER: 25,618
29 REFERENCE/DOCKET NUMBER: HERCEND-2A
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: (202) 628-5197
32 TELEFAX: (202) 737-3528
33 INFORMATION FOR SEQ ID NO: 8:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 457 amino acids
36 TYPE: amino acid
37 STRANDEDNESS: single
38 TOPOLOGY: linear
39 MOLECULE TYPE: protein
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
```

Query Match 43.0%; Score 999; DB 2; Length 457;
Best Local Similarly 52.9%; Pred. No. 2.5e-75;
Matches 207; Conservative 61; Mismatches 115; Indels 8; Gaps 5;

Oy 1 MNRGVFRLH--LVLQLALLPAATQGNKVLGGKKDVTYELTCTASQKKS IQFHWNKNQ 58
||| ||| |::|||: ||| |||||:: ||| :::: |||:::
Db 1 MCRGFSFRHLPLLLQLSKLVVTOGKTVLKGEGGSALPCBESTSRRSASFAMWSSDQ 60

```

Qy 59 IKILNGSGSLTGPSPKSLNDPADRSRLMDQGNFLLIKLIKIDSPYICENVDQKEEV 118
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 KTLIGYXNKULLKIGSELEYSPDSRKNAWERSFPELLINKRMBSQTYTCETENKKEEV 120
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 119 QLVFGLTANSDTHLLQGGSLTLTLES--PGGSPFVOCRSPRGKNIQGGKTLVSOLEQ 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 ELWVFRFTFPGIRLLQGGSLTLILDSNPKSDPPIECKHSSNINVKDSKAFTHSLRIQ 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 178 DSGMTCTVLOQNKVYFKIDIVLAFOKASSIVKKEGEQVESFPFLAFTVEKLITSGGE 237
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 DSGIMNCTVLIQNKHSFPMKLSVIGFASSTITAKSGEASFPFLNLGEEVL--QGE 238
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 238 LMMQABASSSSKSWITFDLKNKESVSKRVTDPKIQMGKULPLHLTLPOALPOYAGSGL 297
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 LRMKAERAPSGQSWITFSLQNKQVASKYSTSNPKQLSETLPLTLQIPVSLQFAGSGL 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 298 TLALAEKGTGLHQBVLVYMRATQLOKQ-LTCEWVGPIPSPKLMLSLKENKEARVGRK 356
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 TLTLD--RGLLYOEVLVYMKVTOQDPSNTLTCEWVGPTSPMRILIKQENOEAARVROEK 356
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 357 PYWVTLNPEAGMMQCLISDGGVLTLESNLIKVL 387
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 VIOVQAEAGVWQCCLISBGEAVKMSKQVL 387
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 63
US-08-630-172-17
: Sequence 17, Application US/08630172
: Patent No. 6060054

APPLICANT: Staerz, Uwe
 TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
 TITLE OF INVENTION: LYMPHOCYTE VETO
 NUMBER OF SEQUENCES: 41
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sheridan Ross & McIntosh
 STREET: 1700 Lincoln Street, 35th Floor
 CITY: Denver
 STATE: Colorado
 COUNTRY: U.S.
 ZIP: 80203
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/630,172
 FILING DATE:
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Connell, Gary J.
 REGISTRATION NUMBER: 32,020
 REFERENCE/DOCKET NUMBER: 2879-36
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (303) 863-9700
 TELEFAX: (303) 863-0223
 INFORMATION FOR SEQ. ID NO. 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 410 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-630-172-17

Query Match	39.2%	Score 912.5	DB 3	Length 410
Best Local Similarity	55.2%	Pred. No. 3.7e-68		
Matches 208; Conservative	27	Mismatches 71		Indels 8

QY 26 NKVVLGKKGDVELTCTASQKKSIOFHWNNOIKILGNQGSFLTKGPSKLNDRADSRRS 85
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MKVLGGKGDPVELTCTASQKKSIOFHWNNOIKILGNQGSFLTKGPSKLNDRADSRRS 60

QY 86 LMDQGNPFLIIKNIKIEDSDTYICEVEDQKEEVOLLVFGLTANSDFHLLQSGSLTTLLES 145
DB 61 LMDQGNPFLIIKNIKIEDSDTYICEVEDQKEEVOLLVFGLTANSDFHLLQSGSLTTLLES 120
QY 146 PGGSSPVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLQNOKKVEFKIDIVLAFAQ 205
DB 121 PGGSSPVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLQNOKKVEFKIDIVLAFAQ 178
QY 206 KASSIYKKEGEVEFSPLAFTVEKLTGSGELMWQAEARASSKSWITFDLKNKEVSVKR 265
DB 179 -----EPRGPTIKPCPCPCAPNLLG-----GPSVFIFPPKIKDVLAMIS 218
QY 266 VT-----QDPKQMG---KKLPHLTLPLALPOYAGSGNLTALAEKTKLHQ 310
DB 219 LSPITVCVVVDSEDDPDVQISWFFVNNVEVH-----TAQTQTHEDY 260
QY 311 EVNLVVRATQLO-----KNLTCEVWGPTSPKMLSLKLENKEAKVSRKRPVWYLP 363
DB 261 NSRLRVVSALPIQHDQMSGKEFKCKV---NNKDLPAPIERTISKPGSVRAPQVYVLP 317
QY 364 EAGMWQ-----CLLSD 374
DB 318 PEEMTKKQVTLTCMVTD 334

RESULT 64
US-09-375-419-17
Sequence 17, Application US/09375419
Patent No. 6264850
GENERAL INFORMATION:
APPLICANT: Staerz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/375,419
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,172
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 410 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-375-419-17

Query Match 39.2%; Score 912.5; DB 3; Length 410;
Best Local Similarity 55.2%; Pred. No. 3,7e-68;
Matches 208; Conservative 27; Mismatches 71; Indels 71; Gaps 8;

QY 26 NKKVLGKKGDTVELTCTASQKKSIOFHWNKSNQIKILGNQGSFLTQGPSKLNDRADSRRS 85

DB 1 NKKVLGKKGDTVELTCTASQKKSIOFHWNKSNQIKILGNQGSFLTQGPSKLNDRADSRRS 60
QY 86 LMDQGNPFLIIKNIKIEDSDTYICEVEDQKEEVOLLVFGLTANSDFHLLQSGSLTTLLES 145
DB 61 LMDQGNPFLIIKNIKIEDSDTYICEVEDQKEEVOLLVFGLTANSDFHLLQSGSLTTLLES 120
QY 146 PGGSSPVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLQNOKKVEFKIDIVLAFAQ 205
DB 121 PGGSSPVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLQNOKKVEFKIDIVLAFAQ 178
QY 206 KASSIYKKEGEVEFSPLAFTVEKLTGSGELMWQAEARASSKSWITFDLKNKEVSVKR 265
DB 179 -----EPRGPTIKPCPCPCAPNLLG-----GPSVFIFPPKIKDVLAMIS 218
QY 266 VT-----QDPKQMG---KKLPHLTLPLALPOYAGSGNLTALAEKTKLHQ 310
DB 219 LSPITVCVVVDSEDDPDVQISWFFVNNVEVH-----TAQTQTHEDY 260
QY 311 EVNLVVRATQLO-----KNLTCEVWGPTSPKMLSLKLENKEAKVSRKRPVWYLP 363
DB 261 NSRLRVVSALPIQHDQMSGKEFKCKV---NNKDLPAPIERTISKPGSVRAPQVYVLP 317
QY 364 EAGMWQ-----CLLSD 374
DB 318 PEEMTKKQVTLTCMVTD 334

RESULT 65
US-08-021-601-12
Sequence 12, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppa, Stephen H.
APPLICANT: Kimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Atora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414, 057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-12

Query Match 39.2%; Score 911; DB 1; Length 903;

Best Local Similarity 100.0%; Pred. No. 1.6e-67;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 KVLGKGGTVELTCTASQKSIQFHWKNSNQIKILGQSGFLTKGPKSLNDRADSRSL 86
DB 727 KVLGKGGTVELTCTASQKSIQFHWKNSNQIKILGQSGFLTKGPKSLNDRADSRSL 786

QY 87 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDPHLLOQGSLLTLTLESP 146
DB 787 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDPHLLOQGSLLTLTLESP 846

QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNOQKVEFKIDIVLA 203
DB 847 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNOQKVEFKIDIVLA 903

RESULT 66
US-08-082-849B-12
; Sequence 12, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Kimpel, Kurt R.
; APPLICANT: Aroza, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; TITLE OF INVENTION: Related Methods
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082,849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 903 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-082-849B-12

Query Match 39.2%; Score 911; DB 1; Length 903;
Best Local Similarity 100.0%; Pred. No. 1.6e-67;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 KVLGKGGTVELTCTASQKSIQFHWKNSNQIKILGQSGFLTKGPKSLNDRADSRSL 86
DB 727 KVLGKGGTVELTCTASQKSIQFHWKNSNQIKILGQSGFLTKGPKSLNDRADSRSL 786

QY 87 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDPHLLOQGSLLTLTLESP 146
DB 847 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDPHLLOQGSLLTLTLESP 146

DB 787 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDPHLLOQGSLLTLTLESP 846
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNOQKVEFKIDIVLA 203
DB 847 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNOQKVEFKIDIVLA 903

RESULT 67
PCT-US94-01624-12
; Sequence 12, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Kimpel, Kurt R.
; APPLICANT: Aroza, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
; STREET: Stewart Street Tower, 20th Floor, One Market
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01624
; FILING DATE: June 25, 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-115
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 903 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-01624-12

Query Match 39.2%; Score 911; DB 5; Length 903;
Best Local Similarity 100.0%; Pred. No. 1.6e-67;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 KVLGKGGTVELTCTASQKSIQFHWKNSNQIKILGQSGFLTKGPKSLNDRADSRSL 86
DB 727 KVLGKGGTVELTCTASQKSIQFHWKNSNQIKILGQSGFLTKGPKSLNDRADSRSL 786

QY 87 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDPHLLOQGSLLTLTLESP 146
DB 787 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDPHLLOQGSLLTLTLESP 846

QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNOQKVEFKIDIVLA 203
DB 847 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNOQKVEFKIDIVLA 903

RESULT 68
US-08-630-172-1
; Sequence 1, Application US/08630172
; Patent No. 6060054
; GENERAL INFORMATION:

APPLICANT: Staerz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
TITLE OF INVENTION: LYMPHOCYTE VETO
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,172
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 138 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-630-172-1

Query Match 30.8%; Score 716; DB 3; Length 138;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 NKVVLGKKGDTVELTCTASQKSIQPHMKNSNOIKILGNQGSFLTKGPSKLNDRADSRRS 85
DB 1 NKVVLGKKGDTVELTCTASQKSIQPHMKNSNOIKILGNQGSFLTKGPSKLNDRADSRRS 60
QY 86 LMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLLOGSILTLTLES 145
DB 61 LMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLLOGSILTLTLES 120
QY 146 PGSSPSVQCRRSPRGKNI 163
DB 121 PGSSPSVQCRRSPRGKNI 138

RESULT 69
US-08-375-419-1
Sequence 1, Application US/09375419
Patent No. 6264950
GENERAL INFORMATION:
APPLICANT: Staerz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
TITLE OF INVENTION: LYMPHOCYTE VETO
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/375,419
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,172
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 138 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-375-419-1

Query Match 30.8%; Score 716; DB 3; Length 138;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 NKVVLGKKGDTVELTCTASQKSIQPHMKNSNOIKILGNQGSFLTKGPSKLNDRADSRRS 85
DB 1 NKVVLGKKGDTVELTCTASQKSIQPHMKNSNOIKILGNQGSFLTKGPSKLNDRADSRRS 60
QY 86 LMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLLOGSILTLTLES 145
DB 61 LMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLLOGSILTLTLES 120
QY 146 PGSSPSVQCRRSPRGKNI 163
DB 121 PGSSPSVQCRRSPRGKNI 138

RESULT 70
US-08-558-269-10
Sequence 10, Application US/08558269
Patent No. 5961973
GENERAL INFORMATION:
APPLICANT: Crea, Roberto
TITLE OF INVENTION: PATHOGEN-TARGETED BIOCATALYSTS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/558,269
FILING DATE: 13-NOV-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/847,800
FILING DATE: 06-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: CRT-001CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-558-269-10

Query Match 25.9%; Score 603; DB 2; Length 376;
Best Local Similarity 46.6%; Pred. No. 2.5e-42;
Matches 159; Conservative 26; Mismatches 96; Indels 60; Gaps 11;

QY 27 KVLKGGKDTVELTCTASQKSIQFHMKNNOIKILGNGSFLLTGPSKLNDRADSRSL 86
DB 3 KVLKGGKDTVELTCTASQKSIQFHMKNNOIKILGNGSFLLTGPSKLNDRADSRSL 62
QY 87 WDQGNFPLIINKLIKIEDSDTYICEVEDQKEVQLVFGLTANSDTHLHGOSLTLLTLESP 146
DB 63 WDQGNFPLIINKLIKIEDSDTYICEVEDQKEVQLVFGLTANSDTHLHGOSLTLLTLESP 122
QY 147 P---GSSP-SVQCSBPRKNIQGGKTLVSQLELDGSGTWCT---VLQNKVVEPKID 198
DB 123 DAELGMSFWQVWLFKSPQELLCGASLI-----SDRWVLTAAHCILYPPMDKNFTEN 174
QY 199 IYVLAFOKASSIVYKKEGEVFEFPLAFTVEKLTGSGELMWQARASSSKSWITFDLKN 258
DB 175 DLVVGKHSRTYERENIEKISM-----LEKTIYHPRYMW---RENLDRIDALMKLK- 223
QY 259 KEVSVK-----RYTQDPKLGKGLPLHLTLFQALPOYAGSGNLTLLAEKATGK- 307
DB 224 KPVAFSDYIHVPCLPDRETAASLLQAGYK-----GRVTGMGNLKETWTANVGKG 272
QY 308 ---LHGEVNLVYMR-----ATQLQKNLTCGEVWGTPSPK 337
DB 273 QPSVLQVYNLPIYERPVCKDSTRIRITDMFCAGYKPDGK 313

RESULT 71

US-09-410-882-10
Sequence 10, Application US/09410882
Patent No. 6287561
GENERAL INFORMATION:
APPLICANT: Crea, Roberto
TITLE OF INVENTION: PATHOGEN-TARGETED BIOCATALYSTS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/410,882
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/558,269
FILING DATE: 13-NOV-1995
APPLICATION NUMBER: US 07/847,800
FILING DATE: 06-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-410-882-10

Query Match 25.9%; Score 603; DB 3; Length 376;
Best Local Similarity 46.6%; Pred. No. 2.5e-42;
Matches 159; Conservative 26; Mismatches 96; Indels 60; Gaps 11;

QY 27 KVLKGGKDTVELTCTASQKSIQFHMKNNOIKILGNGSFLLTGPSKLNDRADSRSL 86
DB 3 KVLKGGKDTVELTCTASQKSIQFHMKNNOIKILGNGSFLLTGPSKLNDRADSRSL 62
QY 87 WDQGNFPLIINKLIKIEDSDTYICEVEDQKEVQLVFGLTANSDTHLHGOSLTLLTLESP 146
DB 63 WDQGNFPLIINKLIKIEDSDTYICEVEDQKEVQLVFGLTANSDTHLHGOSLTLLTLESP 122
QY 147 P---GSSP-SVQCSBPRKNIQGGKTLVSQLELDGSGTWCT---VLQNKVVEPKID 198
DB 123 DAELGMSFWQVWLFKSPQELLCGASLI-----SDRWVLTAAHCILYPPMDKNFTEN 174
QY 199 IYVLAFOKASSIVYKKEGEVFEFPLAFTVEKLTGSGELMWQARASSSKSWITFDLKN 258
DB 175 DLVVGKHSRTYERENIEKISM-----LEKTIYHPRYMW---RENLDRIDALMKLK- 223
QY 259 KEVSVK-----RYTQDPKLGKGLPLHLTLFQALPOYAGSGNLTLLAEKATGK- 307
DB 224 KPVAFSDYIHVPCLPDRETAASLLQAGYK-----GRVTGMGNLKETWTANVGKG 272
QY 308 ---LHGEVNLVYMR-----ATQLQKNLTCGEVWGTPSPK 337
DB 273 QPSVLQVYNLPIYERPVCKDSTRIRITDMFCAGYKPDGK 313

RESULT 72

US-08-558-269-6
Sequence 6, Application US/08558269
Patent No. 5961973
GENERAL INFORMATION:
APPLICANT: Crea, Roberto
TITLE OF INVENTION: PATHOGEN-TARGETED BIOCATALYSTS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/558,269
FILING DATE: 13-NOV-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/847,800
FILING DATE: 06-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 383 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-558-269-6

Query Match 25.4%; Score 589.5; DB 2; Length 383;
Best Local Similarity 72.9%; Pred. No. 3,5e-41;
Matches 121; Conservative 9; Mismatches 23; Indels 13; Gaps 2;

QY 27 KVLGKKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL 86
DB 3 KVLGKKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL 62
QY 87 WDOGNFPLIIKNIKIEDSDTYICEVEDQKEEVOLVFGLTANSDTHLLOGQSLTLLESP 146
DB 63 WDOGNFPLIIKNIKIEDSDTYICEVEDQKEEVOLVFGLTANSDTHLLOGQSLTLLESP 117
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLAQOKK 192
DB 118 --STCGLRQYSQPGFRIKGLFADIA-----SHPQQAIFAKGR 155

RESULT 73
US-09-410-882-6
Sequence 6, Application US/09410882
Patent No. 6287561
GENERAL INFORMATION:
APPLICANT: Crea, Roberto
TITLE OF INVENTION: PATHOGEN-TARGETED BIOCATALYSTS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/410,882
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/558,269
FILING DATE: 13-NOV-1995
APPLICATION NUMBER: US 07/847,800
FILING DATE: 06-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: CRI-001CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 383 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-410-882-6

Query Match 25.4%; Score 589.5; DB 3; Length 383;
Best Local Similarity 72.9%; Pred. No. 3,5e-41;
Matches 121; Conservative 9; Mismatches 23; Indels 13; Gaps 2;
QY 27 KVLGKKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL 86
DB 3 KVLGKKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL 62

DB 3 KVLGKKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL 62
QY 87 WDOGNFPLIIKNIKIEDSDTYICEVEDQKEEVOLVFGLTANSDTHLLOGQSLTLLESP 146
DB 63 WDOGNFPLIIKNIKIEDSDTYICEVEDQKEEVOLVFGLTANSDTHLLOGQSLTLLESP 117
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLAQOKK 192
DB 118 --STCGLRQYSQPGFRIKGLFADIA-----SHPQQAIFAKGR 155

RESULT 74
US-08-466-368-5
Sequence 5, Application US/08466368
Patent No. 6093539
GENERAL INFORMATION:
APPLICANT: Maddon, Paul J.
APPLICANT: Littman, Dan R.
APPLICANT: Chess, Leonard
APPLICANT: Axel, Richard
APPLICANT: Weiss, Robin
APPLICANT: McDougal, J. S.
TITLE OF INVENTION: DNA ENCODING THE T CELL SURFACE PROTEIN
TITLE OF INVENTION: T4 AND USE OF FRAGMENTS OF T4 IN THE TREATMENT OF AIDS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,368
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 24577-EI-B/JPW/ARC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 94 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: YES
FEATURE:
NAME/KEY: Active-site
LOCATION: 1..94
US-08-466-368-5

Query Match 21.2%; Score 494; DB 3; Length 94;
Best Local Similarity 100.0%; Pred. No. 4,2e-34;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 24 QGNKVVLGKKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 83
DB 1 QGNKVVLGKKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 60
QY 84 RSLMDGNFPLIIKNIKIEDSDTYICEVEDQKEE 117
DB 61 RSLMDGNFPLIIKNIKIEDSDTYICEVEDQKEE 94


```
RESULT 75
US-08-470-998-2
; Sequence 2, Application US/08470998
; Patent No. 6570000
; GENERAL INFORMATION:
; APPLICANT: Maddon, Paul J.
; APPLICANT: Littman, Dan R.
; APPLICANT: Chess, Leonard
; APPLICANT: Axel, Richard
; APPLICANT: Weiss, Robin
; APPLICANT: McDougal, J. S.
; TITLE OF INVENTION: DNA ENCODING THE T CELL SURFACE PROTEIN
; TITLE OF INVENTION: T4 AND USE OF FRAGMENTS OF T4 IN THE TREATMENT OF AIDS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,998
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 24577-F1-B/JPW/AKC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 94 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: YES
; FEATURE:
; NAME/KEY: Active-site
; LOCATION: 1..94
; US-08-470-998-2

Query Match      21.2%; Score 494; DB 4; Length 94;
Best Local Similarity 100.0%; Pred. No. 4.2e-34;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      24  QGNKRVLGKGDVLTCTASQKSIQFHWKNSNQIKILNGSFLTGPSKLNDRASR 83
Db      1  QGNKRVLGKGDVLTCTASQKSIQFHWKNSNQIKILNGSFLTGPSKLNDRASR 60

Cy      84  RSLMDQGNFPLIIRLNKIKEDSDTYICEVEDQKEE 117
Db      61  RSLMDQGNFPLIIRLNKIKEDSDTYICEVEDQKEE 94

RESULT 76
US-08-328-500-10
; Sequence 10, Application US/08328500
; Patent No. 6673896
; GENERAL INFORMATION:
; APPLICANT: Maddon, Paul J.
; APPLICANT: Axel, Richard
; APPLICANT: Sweet, Richard W.
```

```
APPLICANT: Arthos, James
; TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,500
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/24577-CY
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 94 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-08-328-500-10

Query Match      21.2%; Score 494; DB 4; Length 94;
Best Local Similarity 100.0%; Pred. No. 4.2e-34;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      24  QGNKRVLGKGDVLTCTASQKSIQFHWKNSNQIKILNGSFLTGPSKLNDRASR 83
Db      1  QGNKRVLGKGDVLTCTASQKSIQFHWKNSNQIKILNGSFLTGPSKLNDRASR 60

Cy      84  RSLMDQGNFPLIIRLNKIKEDSDTYICEVEDQKEE 117
Db      61  RSLMDQGNFPLIIRLNKIKEDSDTYICEVEDQKEE 94

RESULT 77
US-08-332-562A-84
; Sequence 84, Application US/08332562A
; Patent No. 5985599
; GENERAL INFORMATION:
; APPLICANT: MCKENZIE, Ian F.C.
; APPLICANT: HOGARTH, Mark P.
; APPLICANT: HIBBS, Margaret L.
; APPLICANT: SCOTT, Bernadette M.
; APPLICANT: BONADONNA, Lisa
; APPLICANT: HULETT, Mark D.
; TITLE OF INVENTION: FC RECEPTOR FOR IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,562A
; FILING DATE: 31-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/896,457
; FILING DATE: 27-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 54270/119/GRHA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-332-562A-84

```

```

Query Match 17.5%; Score 406; DB 2; Length 80;
Best Local Similarity 98.8%; Pred. No. 7.6e-27;
Matches 79; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 123 FGLTANSDTHLQGSILTLTSPGSSPVQCRSPKXNIQGGKTLVSQLELDGSGTW 182
Db 1 FGLTANSDTHLQGSILTLTSPGSSPVQCRSPKXNIQGGKTLVSQLELDGSGTW 60
Qy 183 TCTVLQNKQKVFEXIDIVVL 202
Db 61 TCTVLQNKQKVFEXIDIVVL 80

```

```

RESULT 78
US-09-050-861B-8
; Sequence 8, Application US/09050861B
; Patent No. 655314
; GENERAL INFORMATION:
; APPLICANT: Payan, Donald
; TITLE OF INVENTION: TOSO AS A TARGET FOR DRUG SCREENING
; FILE REFERENCE: RIGL-002CON
; CURRENT APPLICATION NUMBER: US/09/050,861B
; CURRENT FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: US/09/651,150B
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: US 09/050,861
; PRIOR FILING DATE: 1998-03-30
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-050-861B-8

```

```

Query Match 16.9%; Score 392; DB 4; Length 74;
Best Local Similarity 100.0%; Pred. No. 1e-25;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 37 VELTCTASQKSIQFHMKNNSQIKILGNQGSFLTKGPKSLNDRADSRSLMDQGNPFLII 96
Db 1 VELTCTASQKSIQFHMKNNSQIKILGNQGSFLTKGPKSLNDRADSRSLMDQGNPFLII 60
Qy 97 KNLKIEDSDTYICE 110
Db 61 KNLKIEDSDTYICE 74

```

```

RESULT 79
US-08-332-562A-87

```

```

; Sequence 87, Application US/08332562A
; Patent No. 5985599
; GENERAL INFORMATION:
; APPLICANT: MCKENZIE, Ian F.C.
; APPLICANT: HOGARTH, Mark P.
; APPLICANT: HIBBS, Margaret L.
; APPLICANT: SCOTT, Bernadette M.
; APPLICANT: BONADONNA, Lisa
; APPLICANT: HULETT, Mark D.
; TITLE OF INVENTION: FC RECEPTOR FOR IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,562A
; FILING DATE: 31-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/896,457
; FILING DATE: 27-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 54270/119/GRHA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 83 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-332-562A-87

```

```

Query Match 16.8%; Score 391; DB 2; Length 83;
Best Local Similarity 95.2%; Pred. No. 1.4e-25;
Matches 79; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy 305 TGKLEQVNLVWRATOLQKNTLCEVWGPSPKMLSLKLENKAKVSKREKPVWVNLNPE 364
Db 1 TGKLEQVNLVWRATOLQKNTLCEVWGPSPKMLSLKLENKAKVSKREKPVWVNLNPE 60
Qy 365 AGMWQCLSDSGVLLSNIKVL 387
Db 61 AGMWQCLSDSGVLLSNIKVL 83

```

```

RESULT 80
US-09-100-409A-20
; Sequence 20, Application US/09100409A
; Patent No. 6090388
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
; TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York

```

STATE: NY
COUNTRY: USA
ZIP: 10154-0054
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version
SOFTWARE: #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,409A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 1151-4154
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-751-6849
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-100-409A-20

Query Match 13.5%; Score 314; DB 3; Length 61;
Best Local Similarity 100.0%; Pred. No. 2.5e-19;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 SKLNRAADRSRLWDGNEPLIITKNIKIDSDPTICEVEDDKEEVLVFGILTANSDTHL 133
DB 1 SKLNRAADRSRLWDGNEPLIITKNIKIDSDPTICEVEDDKEEVLVFGILTANSDTHL 60
QY 134 L 134
DB 61 L 61

RESULT 81
US-08-284-391B-35
Sequence 35, Application US/08284391B
Patent No. 5651828
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391B
FILING DATE: 02-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566

FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/565,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 58 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-284-391B-35

Query Match 12.4%; Score 288; DB 2; Length 58;
Best Local Similarity 100.0%; Pred. No. 3.5e-17;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 396 PRASALPAPPTGSALPDPQTASALPDPASALPALALVTSFLGLGVACVLARTR 453
DB 1 PRASALPAPPTGSALPDPQTASALPDPASALPALALVTSFLGLGVACVLARTR 58

RESULT 82
US-09-218-950-35
Sequence 35, Application US/09218950
Patent No. 6284240
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/218,950
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045

TELEX:
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-218-950-35

Query Match 12.4%; Score 288; DB 3; Length 58;
Best Local Similarity 100.0%; Pred. No. 3.5e-17;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 396 PRASALPAPPTGSAIPDPTASALPDPRAASALPALAVISFLIGLGVACVLART 453
Db 1 PRASALPAPPTGSAIPDPTASALPDPRAASALPALAVISFLIGLGVACVLART 58

RESULT 83
US-09-100-409A-10
; Sequence 10, Application US/09100409A
; Patent No. 6090388

GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
; TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND
; TITLE OF INVENTION: IMMUNE DISORDERS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,409A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 1151-4154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-100-409A-10

Query Match 9.8%; Score 228; DB 3; Length 46;
Best Local Similarity 95.6%; Pred. No. 2.6e-12;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 52 HMKNSNQIKILGNQGSFLTKGPKSKLNDRADRSRSLMDQGNFPLII 96
Db 2 HMKNSNQIKILGNQGSFLTKGPKSKLNDRADRSRSLMDQGNFPLII 46

RESULT 84
US-09-100-409A-34
; Sequence 34, Application US/09100409A

Patent No. 6090388
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
; TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND
; TITLE OF INVENTION: IMMUNE DISORDERS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,409A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 1151-4154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 81 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-100-409A-34

Query Match 9.8%; Score 228; DB 3; Length 81;
Best Local Similarity 95.6%; Pred. No. 6.1e-12;
Matches 43; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 52 HMKNSNQIKILGNQGSFLTKGPKSKLNDRADRSRSLMDQGNFPLII 96
Db 37 HMKNSNQIKILGNQGSFLTKGPKSKLNDRADRSRSLMDQGNFPLII 81

RESULT 85
US-09-100-409A-2
; Sequence 2, Application US/09100409A
; Patent No. 6090388
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
; TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND
; TITLE OF INVENTION: IMMUNE DISORDERS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,409A

;; FILING DATE:
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME:
;; REGISTRATION NUMBER:
;; REFERENCE/DOCKET NUMBER: 1151-4154
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 212-758-4800
;; TELEFAX: 212-751-6849
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 40 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-09-100-409A-2

Query Match 9.1%; Score 211; DB 3; Length 40;
Best Local Similarity 97.5%; Pred. No. 5,6e-11;
Matches 39; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 52 HKNSNQIKILNGQSFITKGPSKLNDRADSRSLMDQGN 91
Db 1 HKMNWQIKILNGQSFITKGPSKLNDRADSRSLMDQGN 40

RESULT 86
US-09-100-409A-4
; Sequence 4, Application US/09100409A
; Patent No. 6090388
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
; PREVENTION AND TREATMENT OF HIV INFECTION AND
; TITLE OF INVENTION: IMMUNE DISORDERS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,409A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 1151-4154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-100-409A-4

Query Match 9.1%; Score 211; DB 3; Length 42;
Best Local Similarity 97.5%; Pred. No. 6e-11;
Matches 39; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 52 HKNSNQIKILNGQSFITKGPSKLNDRADSRSLMDQGN 91

Db 2 HKMNWQIKILNGQSFITKGPSKLNDRADSRSLMDQGN 41

RESULT 87
US-09-100-409A-32
; Sequence 32, Application US/09100409A
; Patent No. 6090388
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
; PREVENTION AND TREATMENT OF HIV INFECTION AND
; TITLE OF INVENTION: IMMUNE DISORDERS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,409A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 1151-4154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 77 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-100-409A-32

Query Match 9.1%; Score 211; DB 3; Length 77;
Best Local Similarity 97.5%; Pred. No. 1.5e-10;
Matches 39; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 52 HKNSNQIKILNGQSFITKGPSKLNDRADSRSLMDQGN 91
Db 37 HKMNWQIKILNGQSFITKGPSKLNDRADSRSLMDQGN 76

RESULT 88
5510256-5
; Patent No. 5510256
; APPLICANT: KIRSCHNER, RICHARD J.; MOTT, JOHN E.; ECKENRODE,
; FRANCES M.; BRUNNER, DAVID P.
; TITLE OF INVENTION: ELIMINATING INTERNAL INITIATION OF
; SOLUBLE CDA GENE
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/365,995
; FILING DATE: 29-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 13,828
; FILING DATE: 02-FEB-1993
; APPLICATION NUMBER: 562,861
; FILING DATE: 06-AUG-1990
; SEQ ID NO: 5;
; LENGTH: 41

5510256-5

Query Match 8.6%; Score 200; DB 6; Length 41;
Best Local Similarity 97.6%; Pred. No. 4.8e-10;
Matches 40; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 95 IIKNLIKEDSDTYICEVEDQKEEVQLVGLTANSSTHLLQ 135
Db 1 IIKNLIKEDSDTYICEVEDQKEEVQLVGLTANSSTHLLQ 41

RESULT 89

US-08-630-172-7
; Sequence 7, Application US/08630172
; Patent No. 6060054
; GENERAL INFORMATION:
; APPLICANT: Staerz, Uwe
; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Rose & McIntosh
; STREET: 1700 Lincoln Street, 35th Floor
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630.172
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2879-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-630-172-7

Query Match 7.9%; Score 184.5; DB 3; Length 154;
Best Local Similarity 57.5%; Pred. No. 7e-08;
Matches 46; Conservative 5; Mismatches 12; Indels 17; Gaps 4;

Qy 367 MMOCLSDSG-----QVLLSNI-----KVLPT-----W--STPVHPRASALPAPPTGSA 409
Db 75 MHRQLSDTGTTCQATEENVVYSGTLVLTBEQSGQWHRCSAPPRASALPAPPTGSA 134
Qy 410 LPDPQTASALPDPAPASALP 429
Db 135 LPDPQTASALPDPAPASALP 154

RESULT 90

US-09-375-419-7
; Sequence 7, Application US/09375419
; Patent No. 6264950
; GENERAL INFORMATION:
; APPLICANT: Staerz, Uwe
; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
; NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Rose & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/375.419
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,172
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 154 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-375-419-7

Query Match

7.9%; Score 184.5; DB 3; Length 154;
Best Local Similarity 57.5%; Pred. No. 7e-08;
Matches 46; Conservative 5; Mismatches 12; Indels 17; Gaps 4;

Qy 367 MMOCLSDSG-----QVLLSNI-----KVLPT-----W--STPVHPRASALPAPPTGSA 409
Db 75 MHRQLSDTGTTCQATEENVVYSGTLVLTBEQSGQWHRCSAPPRASALPAPPTGSA 134
Qy 410 LPDPQTASALPDPAPASALP 429
Db 135 LPDPQTASALPDPAPASALP 154

RESULT 91

US-08-332-562A-85
; Sequence 85, Application US/08332562A
; Patent No. 5985599
; GENERAL INFORMATION:
; APPLICANT: MCKENZIE, Ian F.C.
; APPLICANT: HOGARTH, Mark P.
; APPLICANT: HIBBS, Margaret L.
; APPLICANT: SCOTT, Bernadette M.
; APPLICANT: BONADONNA, Lisa
; APPLICANT: HULETT, Mark D.
; TITLE OF INVENTION: FC RECEPTOR FOR IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

RESULT 94
US-09-100-409A-33
Sequence 33, Application US/09100409A
Patent No. 6090388
GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
PREVENTION AND TREATMENT OF HIV INFECTION AND
TITLE OF INVENTION: IMMUNE DISORDERS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10154-0054
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version
SOFTWARE: #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100.409A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 1151-4154
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 69 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-100-409A-33

Query Match 7.2%; Score 167; DB 3; Length 69;
Best Local Similarity 97.0%; Pred. No. 6.1e-07;
Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 64 NQGSFLTKGPKSKLNDRADSRSLMDQGNFPLII 96
Db 37 NQGSFLTKGPKSKLNDRADSRSLMDQGNCPILII 69

RESULT 95
US-09-100-409A-19
Sequence 19, Application US/09100409A
Patent No. 6090388
GENERAL INFORMATION:
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: PEPTIDE COMPOSITION FOR
PREVENTION AND TREATMENT OF HIV INFECTION AND
TITLE OF INVENTION: IMMUNE DISORDERS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10154-0054
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version
SOFTWARE: #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100.409A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 1151-4154
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 29 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-100-409A-19

Query Match 6.7%; Score 156; DB 3; Length 29;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 NQGSFLTKGPKSKLNDRADSRSLMDQGNF 92
Db 1 NQGSFLTKGPKSKLNDRADSRSLMDQGNF 29

RESULT 96
PCT-US93-00031-13
Sequence 13, Application PC/TUS9300031
GENERAL INFORMATION:
APPLICANT: Osborn, Laurelee
APPLICANT: Benjamin, Christopher D.
TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: US
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/00031
FILING DATE: 19930112
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: McNicholas, Janet M.
REGISTRATION NUMBER: 32,918
REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 715-1234
TELEFAX: (312) 715-1234
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-00031-13

Query Match 6.7%; Score 155; DB 5; Length 735;
Best Local Similarity 21.9%; Pred. No. 0.00021;
Matches 91; Conservative 71; Mismatches 166; Indels 88; Gaps 18;


```

QY      0 RHLLVLTQALAPAAI--GQNKVVLGCKDPTVELTCSQKKSIOFHMKNSNKKIILQNG 65
Db      207 RQAKKEIQVTVISPKNTVIVSNPSTKLOEGSGVMTCCSEGLPAPITFW-----254
QY      66 GSFILTKGPKSLANDRADSRSLMDQGNFPLIIKRLIEBSPDYIC---VEDQKEEYOLL 121
Db      255 -----SKKLDNGVLQH---SGNATLTILAMMEBSGIYCEGVNLIGKRKEVELI 303
QY      122 V----FGLTANSOTHLIQ--GOSLTLTLESPPGSSPSVQCSPRKNIQG-----GKT 166
Db      304 VQEKPFVEIVSPGPRIAAOIGDSVMLTCSVMGCESPFSFMRQTIDSPLSGKVRSEGNST 363
QY      169 LVSQLELODSQWTCVTYVLOQNKQKVEFKIDIVYLAFOKASSVYK--KEEYQVPS---222
Db      364 LTLSPAFNEHSHYLTCTVTCGHKKLEKGIQVELYFPPDEPLEMSGGLVNSSTVTSCKV 423
QY      223 ---FPL-AFVVEKLTGSGEIMMOAERASSSKSMITF---DLKNEVSVKKEVTODPKIQ- 273
Db      424 PSYVPLDRLEIELLKKEFTL-----ENIEFLEDTMKLSLEKSLMEITPIIED 472
QY      274 MGKKL-----PLHLTLPOALPOYAGSGNLTALAEATGKLHDEVNLVYMRATQOLK-----N 355
Db      473 TGRALVQOAKLHIDMEFEPEKOROS---TQTLVYVAVP--RDTVLVSPSSIIIEGSSVN 527
QY      326 LTEGWCPTSPKMLSLKLENKEAKSKREKRVWVLNBPAGMOCQLSDSGOYVLE 381
Db      528 MTCISQGFPPKPLKMSRQLPNEGLOPUSENATLTLLISTK-----MEDSGVAYICE 576

```

RESULT 97

PCT-US93-00031-15
; Sequence 15, Application PC/TUS9300031

APPLICANT: Oeshorn, Laurelee
 APPLICANT: Benjamin, Christopher D.
 TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
 TITLE OF INVENTION: IMMUNOGLOBULIN-LIKE DOMAIN OF VCAM1
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Allegraetti & Witcoff, Ltd.
 STREET: 10 South Wacker Drive
 CITY: Chicago
 STATE: IL
 COUNTRY: US
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/00031
 FILING DATE: 19930112
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: McNicholas, Janet M.
 REGISTRATION NUMBER: 32,918
 REFERENCE/DOCKET INFORMATION: 92,306-A; D001 CIP PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 715-1000
 TELEFAX: (312) 715-1234
 INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 716 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein

Query Match	6.7%;	Score 155;	DB 5;	Length 736;
Best Local Similarity	21.9%;	Pred. No. 0.00021;		
Matches	91;	Conservative	71;	Mismatches 166;
			Indels	88;
			Gaps	18

```

Oy      8 RHLLVLTQALAPAA--QGNKVVJGKKDPTVELTCTASQKKSIQFMHKNNOXKILGNQ 65
Db      208 RQAVKELOYVISPKNVIVSNPSTYKLOEGSGVTMCSEGLPAPEITM-----255
Oy      66 GSEFLTGPSPKLNDRADSRSLWDQGNFPIIKNLIKEDSDTYICE---VEDQKEEYVL 121
Db      256 -----SKKDNGNLQHL---SGNATVLTILARMBSGIVYEEGVNLIKRNKEVELI 304
Oy      122 V---FGLTANSDPTHLLQ--GQSLTTLTESPPGSSPVQCSPPGKNIQG-----GKT 166
Db      305 VQEKPEFTVEISPEPRIAQIGDSVMWLTCSVMGCESPSPSMRTQIDSPDGVSSEGTNST 364
Oy      169 LVSQLELODQSGWCTVYQNKXVEFIDIVYLAFOGASSIYVK---KEGEQVEFS---222
Db      365 LTLSPVFSFNEHSHYCTCYVGHKKLEKGIQVELVSPFDPREIEMSGGLVNGSSVTVSCKV 424
Oy      223 ---FPL-AFTVEKTLTSGELMWQAEARASSKSMIWF---DLKNEVSVKRVTDPPKQ- 273
Db      425 PSYVPLDRLEIELLKGFTLL-----ENIEFLEDTMKSLEKSLSEMTPIPIED 473
Oy      274 MGKKL---PLHLTLPQALPOYAGSGNLTALAEATXGKLHQEVLNVYMRATOLQK---N 325
Db      474 TGRALVQAKLTHIDMEFEPPKQROS---TQTLVYNAV--RDTVLVSPSSILIEEGSSVN 528
Oy      326 LTCEWGPSPSPKLMUSLKLENKEAVSRKRPVWVNLNEAGMOCCLSDSQVULLE 381
Db      529 MTLQSGFPPAPKILMSRQLPNEILOPSENNATVLTISYK-----MEDSGVYICE 577

```

RESULT 98

US-08-482-073-6
; Sequence 6, Application US/08482073

```

1 GENERAL INFORMATION:
2 APPLICANT: Heston, Catherine A.
3 APPLICANT: Lobb, Roy R.
4 APPLICANT: Goelz, Susan E.
5 APPLICANT: Osborn, Laurelee
6 APPLICANT: Benjamin, Christopher D.
7 APPLICANT: Rosa, Margaret D.
8 TITLE OF INVENTION: ENDOTHELIAL CELL-LEUKOCYTE ADHESION
9 TITLE OF INVENTION: MOLECULES (SLAMS) AND MOLECULES INVOLVED IN LEUKOCYTE
10 NUMBER OF SEQUENCES: 25
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: Fish & Neave
13 STREET: 1251 Avenue of the Americas
14 CITY: New York
15 STATE: New York
16 COUNTRY: United States of America
17 ZIP: 10020
18
19 COMPUTER READABLE FORM:
20 MEDIUM TYPE: Floppy disk
21 COMPUTER: IBM PC compatible
22 OPERATING SYSTEM: PC-DOS/MS-DOS
23 SOFTWARE: Patentin Release #1.0, Version #1.25
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/08/482,073
26 FILING DATE:
27 CLASSIFICATION:
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: US/08/486,336
30 FILING DATE:
31 APPLICATION NUMBER: US 07/608298
32 FILING DATE: 31-OCT-1990
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: WO PCT/US 90/02357
35 FILING DATE: 27-APR-1990
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER: US 07/452675
38 FILING DATE: 18-DEC-1989
39 PRIOR APPLICATION DATA:
40

```

APPLICATION NUMBER: US 07/359516
 FILING DATE: 01-JUN-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/354151
 FILING DATE: 28-APR-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: Haley Jr., James F.
 REGISTRATION NUMBER: 27,794
 REFERENCE/DOCKET NUMBER: B124CIP4
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 596-9000
 TELEFAX: (212) 596-9090
 TELETYPE: (212) 596-9090
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 739 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-482-073-6

Query Match 6.7%; Score 155; DB 4; Length 739;
 Best Local Similarity 21.9%; Pred. No. 0.00022;
 Matches 91; Conservative 71; Mismatches 166; Indels 88; Gaps 18;

```

Oy 8 RHLLVLQALPLPAT--QGNKVVLGKKGDIVELTCTASQKSIQFHKNSNQIKLGNQ 65
Db 211 RQAVKELOVYISPKRTIVISNPSTKLOBGGSVTMTCSSEGLPAPELFW----- 258
Oy 66 GSFLTKGPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICE---VEDQKEEYQL 121
Db 259 -----SKLDNGNLQHL---SGNATLTLIAMRMEDSGIYCEGVNLIGKRKEVELI 307
Oy 122 V----FGLTANSDTHLQ--GQSLTLTLESPPSSSVQCRSRGKNIQ-----GKT 168
Db 308 VQEKPFVEISPPRIIAQIGDSVWLTCVWGCESSPSFWRTOIDPSLGSKVSEGTNST 367
Oy 169 LSVSOLELQDSGTWTCTVLQNOKVEFKIDIVLAFOKASSIYK---KEGEVFEFS--- 222
Db 368 LTLSPVSENEHSYLTCTYCGHKKLEKIQVELYSFPDPRIEMSGLVNGSSVTVYSCKV 427
Oy 223 ---FPL-AFTVEKLTGSGELMWQERASSSKSWITF---DLNKEVSVKRVTDPKLQ- 273
Db 428 PSVYPLDRLEIELKGETIL-----ENIEFLEDITDKMSLEKSLFETFIPIED 476
Oy 274 MGKTL----PLHLTLQALPQYAGSGNLTLEAKTGLHQBVLVVMRAIQLOK---N 325
Db 477 TGAALVCOAKLHIDMEFEKQROK---TQTLVYNVAP--RDTVLVSPSSILEGSSVN 531
Oy 326 LTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVNLNPEAGMQLSDSGVILE 381
Db 532 MTLCSGFPAPKILMSRQLPNGELQPLSENATLTLISTK-----MEDSGVLYCE 580

```

RESULT 99
 PCT-US93-00031-9
 Sequence 9, Application PC/TUS9300031
 GENERAL INFORMATION:

APPLICANT: Osborn, Laurelee
 APPLICANT: Benjamin, Christopher D.
 TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Allegretti & Witcoff, Ltd.
 STREET: 10 South Wacker Drive
 CITY: Chicago
 STATE: IL
 COUNTRY: US
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/00031
 FILING DATE: 19930112
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: McNicholas, Janet M.
 REGISTRATION NUMBER: 32,918
 REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (312) 715-1000
 TELEFAX: (312) 715-1234
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 739 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US93-00031-9

Query Match 6.7%; Score 155; DB 5; Length 739;
 Best Local Similarity 21.9%; Pred. No. 0.00022;
 Matches 91; Conservative 71; Mismatches 166; Indels 88; Gaps 18;

```

Oy 8 RHLLVLQALPLPAT--QGNKVVLGKKGDIVELTCTASQKSIQFHKNSNQIKLGNQ 65
Db 211 RQAVKELOVYISPKRTIVISNPSTKLOBGGSVTMTCSSEGLPAPELFW----- 258
Oy 66 GSFLTKGPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICE---VEDQKEEYQL 121
Db 259 -----SKLDNGNLQHL---SGNATLTLIAMRMEDSGIYCEGVNLIGKRKEVELI 307
Oy 122 V----FGLTANSDTHLQ--GQSLTLTLESPPSSSVQCRSRGKNIQ-----GKT 168
Db 308 VQEKPFVEISPPRIIAQIGDSVWLTCVWGCESSPSFWRTOIDPSLGSKVSEGTNST 367
Oy 169 LSVSOLELQDSGTWTCTVLQNOKVEFKIDIVLAFOKASSIYK---KEGEVFEFS--- 222
Db 368 LTLSPVSENEHSYLTCTYCGHKKLEKIQVELYSFPDPRIEMSGLVNGSSVTVYSCKV 427
Oy 223 ---FPL-AFTVEKLTGSGELMWQERASSSKSWITF---DLNKEVSVKRVTDPKLQ- 273
Db 428 PSVYPLDRLEIELKGETIL-----ENIEFLEDITDKMSLEKSLFETFIPIED 476
Oy 274 MGKTL----PLHLTLQALPQYAGSGNLTLEAKTGLHQBVLVVMRAIQLOK---N 325
Db 477 TGAALVCOAKLHIDMEFEKQROK---TQTLVYNVAP--RDTVLVSPSSILEGSSVN 531
Oy 326 LTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVNLNPEAGMQLSDSGVILE 381
Db 532 MTLCSGFPAPKILMSRQLPNGELQPLSENATLTLISTK-----MEDSGVLYCE 580

```

RESULT 100
 US-08-261-304-7
 Sequence 7, Application US/08261304
 Patent No. 5708147
 GENERAL INFORMATION:

APPLICANT: Cybulsky, Myron I.
 APPLICANT: Gimbrone, Michael A.
 TITLE OF INVENTION: Mononuclear Leukocyte Directed
 TITLE OF INVENTION: Endothelial Adhesion Molecule Associated with
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 STREET: 1225 Connecticut Avenue, N.W.
 CITY: Suite 300
 STATE: Washington
 COUNTRY: United States of America

```

1      ZIP: 20036
2      COMPUTER READABLE FORM:
3      MEDIUM TYPE: Floppy Disk
4      COMPUTER: IBM PC compatible
5      OPERATING SYSTEM: PC-DOS/MS-DOS
6      SOFTWARE: Ascii
7      CURRENT APPLICATION DATA:
8      APPLICATION NUMBER: US/08/261,304
9      FILING DATE:
10     CLASSIFICATION: 435
11     PRIOR APPLICATION DATA:
12     APPLICATION NUMBER: 07/649,565
13     FILING DATE: 01-FEB-1991
14     APPLICATION NUMBER: U.S. 07/487,038
15     FILING DATE: 02-MAR-1990
16     ATTORNEY/AGENT INFORMATION:
17     NAME: Potter, Jane E. R.
18     REGISTRATION NUMBER: 33,332
19     REFERENCE/DOCKET NUMBER: 0627.2100004
20     TELECOMMUNICATION INFORMATION:
21     TELEPHONE: (202) 833-7533
22     INFORMATION FOR SEQ ID NO. 7:
23     SEQUENCE CHARACTERISTICS:
24     LENGTH: 662 amino acids
25     TYPE: amino acid
26     TOPOLOGY: linear
27     MOLECULE TYPE: peptide
28     US-08-261-304-7

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 13:00:24 ; Search time 10.4109 Seconds
(without alignments)
4185.504 Million cell updates/sec

Title: SEQ5
Perfect score: 2325
Sequence: 1 NMRGVFRRLLVLQLALP.....VISFLGLGLGVACVLRNR 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 125 summaries

Database :
1: PIR 78:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2032	87.4	458	1	RWHT4	T-cell surface gly
2	1884	81.0	432	1	RMCT4	T-cell surface gly
3	1743	75.0	432	2	RMCT4	T-cell surface gly
4	1147	49.3	459	2	A46254	CD4 precursor - ra
5	1101	47.4	432	2	S30193	T-cell surface gly
6	999	43.0	457	2	A27449	T-cell surface gly
7	993	42.7	457	1	RMST4	T-cell surface gly
8	357	15.4	71	2	I60082	CD4 receptor - hum
9	305.5	13.1	99	2	S21461	T-cell surface gly
10	293	12.6	240	2	A39016	T-cell surface gly
11	280.5	12.1	99	2	S21462	T-cell surface gly
12	162	7.0	739	2	JS0675	vascular cell adhe
13	155.5	6.7	538	2	UC2457	vascular cell adhe
14	145	6.7	739	2	A41288	vascular cell adhe
15	144	6.2	739	1	IJHUNG	vascular cell adhe
16	143	6.2	739	2	JN0581	vascular cell adhe
17	143	6.2	647	2	B41288	vascular cell adhe
18	138	5.9	398	2	I49443	gene 28A protein -
19	137.5	5.9	702	2	A36319	carcinoembryonic a
20	135	5.8	1259	2	S36126	neural cell adhesi
21	134.5	5.8	1260	1	S05479	neural cell adhesi
22	133	5.7	2783	2	T34416	hypothetical prote
23	132	5.7	1011	2	T13669	neuromusculin - fr
24	132	5.7	6831	2	A88852	protein unc-22 [im
25	132	5.7	6839	2	S57242	twitelin [similar]
26	132	5.7	1260	2	T27935	hypothetical prote
27	131	5.6	7157	1	A41060	neural cell adhesi
28	131	5.6	5175	2	T20992	hypothetical prote
29	131	5.6	5198	2	T43290	hemiscentin precure

30	130.5	5.6	304	2	S04663	T-cell receptor ga
31	130	5.6	1906	1	S68235	myosin-light-chain
32	130	5.6	4162	2	T42633	connectin/citin -
33	127	5.5	2629	2	T32735	telomerase-associ
34	127	5.5	3707	2	S18252	heparan sulfate pr
35	125.5	5.4	725	1	IJMSNG	neural cell adhesi
36	125.5	5.4	1447	2	A54100	tumor suppressor p
37	125	5.4	1496	2	T08851	Down syndrome cell
38	125	5.4	6805	2	S20901	titin - rabdic (fr
39	124	5.3	1115	1	IJMSNL	neural cell adhesi
40	124	5.3	1298	2	A64157	hypothetical prote
41	124	5.3	7562	2	I38346	elastic citlin - hu
42	123.5	5.3	323	2	S01895	T-cell receptor ga
43	123.5	5.3	1091	1	IJCHNL	neural cell adhesi
44	122.5	5.3	4391	2	A38096	perlecan precursor
45	122	5.2	1323	2	PN0568	connectin 3B - chl
46	121.5	5.2	338	2	JC4776	limbic-system-asso
47	121	5.2	946	2	S28061	SCP1 protein - rat
48	121	5.2	1427	2	I51669	tumor suppressor -
49	120.5	5.2	333	2	A31923	amalgam protein pr
50	120.5	5.2	773	1	QRRBG	secretory componen
51	120.5	5.2	3305	2	T18358	apolipoprotein prec
52	119	5.1	584	2	T08678	hypothetical prote
53	118	5.1	548	2	JC4917	signal transducing
54	118	5.1	6642	2	T29757	protein UNC-89 - C
55	117.5	5.1	858	1	IJRTNC	neural cell adhesi
56	117.5	5.1	1197	2	T30581	neural cell adhesi
57	117.5	5.1	1232	2	T43027	neural cell adhesi
58	116.5	5.0	1339	1	A32579	neuroglial - fruit
59	116.5	5.0	1330	2	S49010	embryonic receptor
60	116	5.0	458	2	JC1509	biliary glycoprote
61	116	5.0	521	2	S34338	neural cell adhesi
62	116	5.0	853	1	IJBONC	neural cell adhesi
63	116	5.0	1021	1	I39207	leukocyte surface
64	115.5	5.0	1136	1	S57845	protein-tyrosine k
65	115.5	4.9	554	2	A69392	hypothetical prote
66	115	4.9	26926	1	I38344	titin, cardiac mus
67	114	4.9	257	2	S00682	IGF Fc receptor al
68	114	4.9	765	2	C42632	cell adhesion mole
69	114	4.9	812	2	B42632	cell adhesion mole
70	114	4.9	932	2	A42632	cell adhesion mole
71	113.5	4.9	245	2	A30154	IGF receptor alpha
72	113.5	4.9	338	2	JC5519	50K glycoprotein p
73	113.5	4.9	1273	2	T42405	sax-3 protein - Ca
74	113.5	4.9	1277	2	T30532	neural cell adhesi
75	113.5	4.9	1849	2	C41859	IGF-specific metal
76	113.5	4.9	2222	2	T13924	sdk protein - fru1
77	113	4.9	540	2	JC4916	signal transducing
78	113	4.9	729	2	AC3616	succinoglycan bios
79	113	4.9	1028	2	I58164	BIG-1 protein - ra
80	113	4.9	1333	2	I78875	receptor tyrosine
81	113	4.9	2130	2	AB0821	probable exported
82	112	4.8	588	2	JH0506	adhesion molecule
83	112	4.8	1091	2	A58532	glial cell membran
84	111.5	4.8	120	2	S46374	Ig kappa chain V-J
85	111.5	4.8	422	2	I37891	interleukin-11 rec
86	111.5	4.8	750	2	S41051	fibroblast growth
87	111	4.8	340	2	S03517	T-cell receptor ga
88	111	4.8	514	2	A44100	cell adhesion mole
89	111	4.8	588	2	A45254	surface glycoprote
90	110.5	4.8	279	2	C81412	NOG1/NO2/sun faml
91	110.5	4.8	1023	2	AE1280	ATP-dependent dBDN
92	109.5	4.7	584	2	I50419	8-glycerin precuro
93	109	4.7	1028	2	A53449	plasmacytoma-asso
94	109	4.7	1367	2	A41228	protein-tyrosine k
95	108	4.6	523	2	I50478	neuroilin - goldfis
96	108	4.6	538	2	I68093	PRR2 delta - human
97	107.5	4.6	122	2	S40370	Ig kappa chain - h
98	107.5	4.6	458	2	S68177	C-CM2a protein is
99	107.5	4.6	519	2	A44063	ecto-ATPase precu
100	107.5	4.6	1138	1	S24066	protein-tyrosine k
101	107.5	4.6	1443	2	I50600	neogenin - chicken
102	107.5	4.6	2629	2	T30987	telomerase-associ

T-cell receptor ga
myosin-light-chain
connectin/citin -
telomerase-associ
heparan sulfate pr
neural cell adhesi
tumor suppressor p
Down syndrome cell
titin - rabdic (fr
neural cell adhesi
hypothetical prote
elastic citlin - hu
T-cell receptor ga
neural cell adhesi
perlecan precursor
connectin 3B - chl
limbic-system-asso
SCP1 protein - rat
tumor suppressor -
amalgam protein pr
secretory componen
apolipoprotein prec
hypothetical prote
signal transducing
protein UNC-89 - C
neural cell adhesi
neural cell adhesi
neural cell adhesi
neuroglial - fruit
embryonic receptor
biliary glycoprote
neural cell adhesi
neural cell adhesi
leukocyte surface
protein-tyrosine k
hypothetical prote
titin, cardiac mus
IGF Fc receptor al
cell adhesion mole
cell adhesion mole
cell adhesion mole
IGF receptor alpha
50K glycoprotein p
sax-3 protein - Ca
neural cell adhesi
IGF-specific metal
sdk protein - fru1
signal transducing
succinoglycan bios
BIG-1 protein - ra
receptor tyrosine
probable exported
adhesion molecule
glial cell membran
Ig kappa chain V-J
interleukin-11 rec
fibroblast growth
T-cell receptor ga
cell adhesion mole
surface glycoprote
NOG1/NO2/sun faml
ATP-dependent dBDN
8-glycerin precuro
plasmacytoma-asso
protein-tyrosine k
neuroilin - goldfis
PRR2 delta - human
Ig kappa chain - h
C-CM2a protein is
ecto-ATPase precu
protein-tyrosine k
neogenin - chicken
telomerase-associ

103	107	4.6	458	1	MMMSR1
104	107	4.6	514	2	A31643
105	107	4.6	521	2	JC1508
106	106.5	4.6	111	2	B37266
107	106.5	4.6	111	2	I38740
108	106.5	4.6	748	2	S41050
109	106	4.6	345	2	S03199
110	106	4.6	345	2	A46052
111	106	4.6	577	2	I50731
112	106	4.6	1018	2	JC4211
113	106	4.6	1070	2	T25836
114	105.5	4.5	293	2	S25741
115	105.5	4.5	538	2	AB1236
116	105.5	4.5	1150	2	B47114
117	105	4.5	345	2	JC4025
118	104.5	4.5	458	2	S23969
119	104.5	4.5	515	1	VCJLGA
120	104	4.5	437	2	A54595
121	104	4.5	528	2	B75364
122	104	4.5	862	2	I49583
123	104	4.5	1017	2	PC4035
124	104	4.5	1259	2	A43425
125	104	4.5	1268	1	A39640

ALIGNMENTS

RESULT 1

RNMUT4

T-cell surface glycoprotein CD4 precursor [validated] - human

N.Alternate names: T-cell surface antigen T4/Leu 3

C.Species: Homo sapiens (man)

C.Date: 28-May-1986 #sequence revision 31-Dec-1988 #text change 20-Apr-2001

C.Accession: A90872; A32722; A34194; A53287; I54176; I54297; A02109; A30039

R.Maddon, P.J.; Littman, D.R.; Godfrey, M.; Maddon, D.R.; Chess, L.; Axel, R.

Cell 42, 93-104, 1985

A.Title: The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein

A.Reference number: A90872; MUID:8554948; PMID:2990730

A.Accession: A90872

A.Molecule type: mRNA

A.Residues: 1-25, 'N', 27-458 <MAD>

A.Experimental source: clone pT48

R.Littman, D.R.; Maddon, P.J.; Axel, R.

Cell 55, 541, 1988

A.Title: Corrected CD4 sequence.

A.Reference number: A90907; MUID:89028665; PMID:3263213

A.Contents: annotation; revision to residue 26

R.Camerini, D.; Seed, B.

Cell 60, 747-754, 1990

A.Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the variable region

A.Reference number: A32722; MUID:90182664; PMID:2107024

A.Accession: A32722

A.Stature: nucleic acid sequence not shown; not compared with conceptual translation

A.Molecule type: mRNA

A.Residues: 26-426, 428-458 <CAM>

R.Carr, S.A.; Hemling, M.E.; Folena-Wasserman, G.; Sweet, R.W.; Anumula, K.; Barr, J.R.; J. Biol. Chem. 264, 21286-21295, 1989

A.Title: Protein and carbohydrate structural analysis of a recombinant soluble CD4 receptor

A.Reference number: A34194; MUID:90078332; PMID:2592374

A.Contents: disulfide bonds; carbohydrate-binding sites

A.Accession: A34194

A.Molecule type: protein

A.Residues: 26-394 <CAR>

R.Lederman, S.; DeMartino, J.A.; Daugherty, B.L.; Foeldvari, I.; Yellin, M.J.; Cleary, A.

Mol. Immunol. 28, 1171-1181, 1991

A.Title: A single amino acid substitution in a common African allele of the CD4 molecule

A.Reference number: A53287; MUID:92072595; PMID:1961196

A.Accession: A53287

A.Status: not compared with conceptual translation

A.Molecule type: mRNA

A.Residues: 250-264, 'W', 266-280 <LED>

A.Note: sequence extracted from NCBI backbone (NCBIP:68249)

biliary glycoprote
cell adhesion 80K
biliary glycoprote
Ig kappa chain V r
Ig kappa chain V r
fibroblast growth
opoid-binding pro
vascular cell adhe
Ig heavy chain - n
neural adhesion pr
hypothetical prote
Ig lambda chain -
interleukin protein
phosphoprotein pho
opoid-binding cel
cell-adhesion mole
env polypeptide pr
transcription fact
extracellular solu
differentiation an
cell-cycle-depend
Bravo/Nr-CAM cell
neural cell adhe

R.Edwards, M.C.; Gibbs, R.A.
Genomics 14, 590-597, 1992
A.Title: A human dimorphism resulting from loss of an Alu.
A.Reference number: I54176; MUID:93052387; PMID:1330888
A.Accession: I54176
A.Stature: translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-72 <RES>
A.Cross-references: GB:U47924; GB:M86525; GB:U72506; NID:91633547; PIDN:AA51309.1; PID:
R.Hodge, T.W.; Saaso, D.R.; McDougal, J.S.
Hum. Immunol. 30, 99-104, 1991
A.Title: Humans with OKT4-epitope deficiency have a single nucleotide base change in the
A.Reference number: I54297; MUID:91216786; PMID:1708753
A.Accession: I54297
A.Stature: translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-264, 'W', 266-458 <RE2>
A.Cross-references: GB:M5160; NID:9179143; PIDN:AAA16069.1; PID:9179144
C.Comment: Macrophage tropic strains of HIV-1 bind to a complex of chemokine (C-C) recep
C.Genetics:
A.Gene: GDB:CD4
A.Cross-references: GDB:119767; OMIM:186940
A.Map position: 12pter-12p12
A.Introns: 16/3
C.Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C.Keywords: AIDS; duplication; glycoprotein; T-cell; transmembrane protein
F.1-25/Domain: signal sequence #status predicted <SIG>
F.26-458/Product: T-cell surface glycoprotein CD4 #status experimental <MAT>
F.134-111/Domain: immunoglobulin homology <IM1>
F.136-166/Domain: immunoglobulin homology <IM2>
F.216-299/Domain: immunoglobulin homology <IM3>
F.321-372/Domain: immunoglobulin homology <IM4>
F.397-420/Domain: transmembrane #status predicted <TM>
F.421-458/Domain: intracellular #status predicted <INT>
F.41-109, 155-184, 328-370/Disulfide bonds: #status experimental
F.296, 325/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 87.4%; Score 2032; DB 1; Length 458;
Best Local Similarity 89.2%; Pred. No. 2, 1e-127;
Matches 405; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

QY	1	MNRGVPFRHLILVQLALPPATQGNKRVLGKGDVLELTCTASQKSIQFMKNSQIK	60
DB	1	MNRGVPFRHLILVQLALPPATQGNKRVLGKGDVLELTCTASQKSIQFMKNSQIK	60
QY	61	ILNGQSFLLTKGPKLNDRADSRSLMDQGNFPLIRNLKIEDSDTYICEVEDQKEVQL	120
DB	61	ILNGQSFLLTKGPKLNDRADSRSLMDQGNFPLIRNLKIEDSDTYICEVEDQKEVQL	120
QY	121	LVFGLTNSDTHLQGSLLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG	180
DB	121	LVFGLTNSDTHLQGSLLTTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG	180
QY	181	TWTCVILONQKVEFKIDIVLAFQKASSIYKKEGSEVERSPFLAFTVEKLTGSGELMW	240
DB	181	TWTCVILONQKVEFKIDIVLAFQKASSIYKKEGSEVERSPFLAFTVEKLTGSGELMW	240
QY	241	QAEASSSKSWITTDLNKKEYSVKRYQDPQLQNGKPLHLTLTPQALPOVAGSGNLTLA	300
DB	241	QAEASSSKSWITTDLNKKEYSVKRYQDPQLQNGKPLHLTLTPQALPOVAGSGNLTLA	300
QY	301	LEAKTGKLGHOVNLVWRATQLOKNTLCEVWGPSPKMLSTLKENKAKSKKEKPYVW	360
DB	301	LEAKTGKLGHOVNLVWRATQLOKNTLCEVWGPSPKMLSTLKENKAKSKKEKPYVW	360
QY	361	LNPRAGMWQCLSDSGVLTESNISKVLPTWSTPVHPRASALPAPPTGSLDPPQTASALP	420
DB	361	LNPRAGMWQCLSDSGVLTESNISKVLPTWSTPVHPRASALPAPPTGSLDPPQTASALP	420
QY	421	DPPASALPALAYISFLIGLGV-ACVLAATR	453
DB	397	-----MALIVLGVAAGLLFLGLGIFPCVRCRHR	425

RESULT 2

RMCT4

T-cell surface glycoprotein CD4 - chimpanzee

N/Alternate names: T-cell surface antigen T4/Leu 3

C/Species: Pan troglodytes (chimpanzee)

C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999

C/Accession: B32722; A46534

R/Camerini, D.; Seed, B.

Cell 60, 747-754, 1990

A/Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the vi

A/Reference number: A32722; MUID:90182664; PMID:2107024

A/Accession: B32722

A/Molecule type: mRNA

A/Residues: 1-432 <CAM>

A/Cross-references: GB:M31135

R/Fomsgaard, A.; Hirsch, V.M.; Johnson, P.R.

Eur. J. Immunol. 22, 2973-2981, 1992

A/Title: Cloning and sequences of primate CD4 molecules: diversity of the cellular recep

A/Reference number: A46534; MUID:93049640; PMID:1425921

A/Accession: A46534

A/Status: not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 3-399 <FOM>

A/Note: sequence extracted from NCBI backbone (NCBIP:118332)

C/Comment: This protein is expressed on most thymocytes, as a subset of mature T-cells

C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology

C/Keywords: duplication; glycoprotein; T-cell; transmembrane protein

F/1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>

F/1-371/Domains: extracellular #status predicted <EXT>

F/9-86/Domains: immunoglobulin homology <IM1>

F/111-161/Domains: immunoglobulin homology #status atypical <IM2>

F/191-274/Domains: immunoglobulin homology <IM3>

F/396-347/Domains: immunoglobulin homology <IM4>

F/372-395/Domains: transmembrane #status predicted <TM>

F/396-432/Domains: intracellular #status predicted <INT>

F/16-84,130-159,303-345/Disulfide bonds: #status predicted

F/271,300/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 81.0%; Score 1884; DB 1; Length 432;

Best Local Similarity 87.9%; Pred. No. 1,3e-117;

Matches 346; Conservative 2; Mismatches 20; Indels 30; Gaps 2;

27 KYVLGGKGVTEVLTCTASQKKSIOFHWKNSQIKILGNGSFLTKGPSKLNDRADSRSL 86

2 KYVLGGKGVTEVLTCTASQKKSIOFHWKNSQIKILGNGSFLTKGPSKLNDRADSRSL 61

87 WDOGNEPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSDFHLLQGSLTTLTLESP 146

62 WDOGNEPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSDFHLLQGSLTTLTLESP 121

147 PGSSPVQCRSPGKNIQGGKTLVSQLELDSDGTCTVLONQKVEKIDIVLAFQK 206

122 PGSSPVQCRSPGKNIQGGKTLVSQLELDSDGTCTVLONQKVEKIDIVLAFQK 181

207 ASSIVYKKEGEVEFSPFLAFVTEKLTGSGELMWOAERASSKSWTTPDLKNKEVSVKRV 266

182 ASSIVYKKEGEVEFSPFLAFVTEKLTGSGELMWOAERASSKSWTTPDLKNKEVSVKRV 241

267 TDDPKLQMGKKLPLHLTPALPOYAGSGNLTALAEAKTGKLGHOEVNLMVMAATQLOK 326

242 TDDPKLQMGKKLPLHLTPALPOYAGSGNLTALAEAKTGKLGHOEVNLMVMAATQLOK 301

327 TCEVWGPTSPKMLSLKLENKAQVSKREKPVVNLNPEAGMOCCLSDSGVLLBSNIVY 386

302 TCEVWGPTSPKMLSLKLENKAQVSKREKPVVNLNPEAGMOCCLSDSGVLLBSNIVY 361

387 LPTWSPVHPRASALPAPPTGSALPDPOFASALPDPPASALPALAVISFLIGLGLGV- 445

362 LPTWSPVHP- - - - -MALIVLGVAAGLLFTGLGLF 392

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

Db 393 FCVRCRHR 400

RESULT 3

RMWOT4

T-cell surface glycoprotein CD4 - rhesus macaque

N/Alternate names: T-cell surface antigen T4/Leu 3

C/Species: Macaca mulatta (rhesus macaque)

C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999

C/Accession: C32722

R/Camerini, D.; Seed, B.

Cell 60, 747-754, 1990

A/Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the vi

A/Reference number: A32722; MUID:90182664; PMID:2107024

A/Accession: C32722

A/Molecule type: mRNA

A/Residues: 1-432 <CAM>

A/Cross-references: GB:M31134

C/Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells

C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology

C/Keywords: duplication; glycoprotein; T-cell; transmembrane protein

F/1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>

F/1-371/Domains: extracellular #status predicted <EXT>

F/9-86/Domains: immunoglobulin homology <IM1>

F/111-161/Domains: immunoglobulin homology #status atypical <IM2>

F/180-293/Domains: immunoglobulin homology <IM3>

F/396-347/Domains: immunoglobulin homology <IM4>

F/372-395/Domains: transmembrane #status predicted <TM>

F/396-432/Domains: intracellular #status predicted <INT>

F/16-84,130-159,303-345/Disulfide bonds: #status predicted

F/271,300/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 75.0%; Score 1743; DB 1; Length 432;

Best Local Similarity 81.1%; Pred. No. 2,9e-108;

Matches 347; Conservative 17; Mismatches 34; Indels 30; Gaps 2;

27 KYVLGGKGVTEVLTCTASQKKSIOFHWKNSQIKILGNGSFLTKGPSKLNDRADSRSL 86

2 KYVLGGKGVTEVLTCTASQKKSIOFHWKNSQIKILGNGSFLTKGPSKLNDRADSRSL 61

87 WDOGNEPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSDFHLLQGSLTTLTLESP 146

62 WDOGNEPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSDFHLLQGSLTTLTLESP 121

147 PGSSPVQCRSPGKNIQGGKTLVSQLELDSDGTCTVLONQKVEKIDIVLAFQK 206

122 PGSSPVQCRSPGKNIQGGKTLVSQLELDSDGTCTVLONQKVEKIDIVLAFQK 181

207 ASSIVYKKEGEVEFSPFLAFVTEKLTGSGELMWOAERASSKSWTTPDLKNKEVSVKRV 266

182 ASSIVYKKEGEVEFSPFLAFVTEKLTGSGELMWOAERASSKSWTTPDLKNKEVSVKRV 241

267 TDDPKLQMGKKLPLHLTPALPOYAGSGNLTALAEAKTGKLGHOEVNLMVMAATQLOK 326

242 TDDPKLQMGKKLPLHLTPALPOYAGSGNLTALAEAKTGKLGHOEVNLMVMAATQLOK 301

327 TCEVWGPTSPKMLSLKLENKAQVSKREKPVVNLNPEAGMOCCLSDSGVLLBSNIVY 386

302 TCEVWGPTSPKMLSLKLENKAQVSKREKPVVNLNPEAGMOCCLSDSGVLLBSNIVY 361

387 LPTWSPVHPRASALPAPPTGSALPDPOFASALPDPPASALPALAVISFLIGLGLGV- 445

362 LPTWSPVHP- - - - -MALIVLGVAAGLLFTGLGLF 392

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

446 ACVLARTR 453

```

CjDate: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jan-2000
CjAccession: A46254
RjHague, B.F.; Sawesdikosol, S.; Brown, T.J.; Lee, K.; Recker, D.P.; Kindt, T.J.
Proc Natl. Acad. Sci. U.S.A. 89, 7963-7967, 1992
AjTitle: CD4 and its role in infection of rabbit cell lines by human immunodeficiency vi
AjReference number: A46254; MUID:92390370; PMID:1518821
AjAccession: A46254
AjStatus: preliminary
AjMolecule type: mRNA
AjResidues: 1-459 <HAG>
AjCross-references: GB:M92840; NID:9164871; PIDN:AAA1198.1; PID:9164872
AjNote: sequence extracted from NCBI backbone (NCBIN:112732, NCBI:112733)
CjSuperfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
Fj322-372/Domain: immunoglobulin homology <IMM>

Query Match      49.3%; Score 1147; DB 2; Length 459;
Best Local Similarity 58.0%; Pred. No. 1,1e-68;
Matches 240; Conservative 68; Mismatches 94; Indels 12; Gaps 6;

QY 1 MNRGVPRHL-LIVLQALIPATQGNKVLGKGGDTVELTCTASQKSIQFHWKNSNQ 60
DB 1 MNRRIYFQCLLVLPALIPATGWKTIVRGKAGALVELPCOSSQKNSVFWMKHAQVK 60
QY 61 ILGNQGS---SFLTGKPSKLNDRADSRSLMDQGNPPLTIKUKIEDSDTYICEVDQKE 116
DB 61 ILGNQGSSSSSFWLKGNSPLSNRVESKKMMQDGSFPLVTKDRLMDSDGTIVCEVDKKM 120
QY 117 EVQLLVFGITANSPTHLLOGQSITLTLESPPGSSPSVQCRSPRGKNIQSGKTLVSQDEL 176
DB 121 EVELLVFRITANPNRRLHLHGQSITLTLEBSPVSGSPVQKSPENKIIIEGPIPCSPKRL 180
QY 177 QDSGTWCTV-LQNKQVEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGS 235
DB 181 QDSGTWCHLSFQDQNKELDIKITVIGPRKASATVYKKEGQVFSPLNFEDSL--S 238
QY 236 GELMQAERASSSKSWITFDLKNKESVYKRVTDPELQMKKLPHLTPQALPYAGSG 295
DB 239 GELMQQVDASSAQSWSPSLDRKQSVQKIIIPDKIQMSKPLPLSTLPQALHRYASG 298
QY 236 NLTLEAKTGLHDEVNLVVMRATQLOKNTLCEWGPSPPTLSIKLENKEAVSKRE 355
DB 239 NLTSLTD--KGLKHQQSLVLMKVQVKNKLTCEVGLPDPKMLSLKLEDEAKVS--TQ 355
QY 356 KEVWVLPNPEAGMOCILSDSGVLTLESNIKVLPTWSTVPVPRASALPAPTGSA 409
DB 356 KKVQVLPDPRAGTWQCCLSGDVLLESKADVALATGS--HQRPITLALGSGTA 407

RESULT 5
S30193
T-cell surface glycoprotein CD4 - dog
CjSpecies: Canis lupus familiaris (dog)
CjDate: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jul-2000
CjAccession: S30193
RjMilde, K.F.; Conner, G.E.; Mintz, D.H.; Alejandro, R.
Biochim. Biophys. Acta 1172, 315-318, 1993
AjTitle: Primary structure of the canine CD4 antigen.
AjReference number: S30193; MUID:93192324; PMID:7916632
AjAccession: S30193
AjStatus: preliminary
AjMolecule type: mRNA
AjResidues: 1-432 <ML>
AjCross-references: EMBL:X68865; NID:9288652; PIDN:CB37664.1; PID:94467377
CjSuperfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
Fj202-311/Domain: immunoglobulin homology <IMM>

Query Match      47.4%; Score 1101; DB 2; Length 432;
Best Local Similarity 52.2%; Pred. No. 1.2e-65;
Matches 235; Conservative 66; Mismatches 97; Indels 52; Gaps 8;

QY 12 LVQLALIPATQGNKVLGKGGDTVELTCTASQKSIQFHWKNSNQIKILGNQGSFLTK 71
DB 12 LVQLALIPATQGNKVLGKGGDTVELTCTASQKSIQFHWKNSNQIKILGNQGSFLTK 71

```

```

DB 1 LMLQLVMLPAVTPVREVLVGRAGDAVELPCOTSQKNIHFNMWRDSMWQILGNQGSFWTV 60
QY 72 GSKLNDRADSRRLMPOGNPPLTIKUKIEDSDTYICEVDQKEVQLVFGTLA----- 127
DB 61 GSKRLKRVESKSLMQDGSFPLVTKDEVADSIGTIFCDT-DKQVEVLVFNLTAKKDS 119
QY 128 -----NSDTHLLOGQSITLTLESPPGSSPSVQCRSPRGKNIQSGKTLVSQLELDQSGTW 182
DB 120 GSSGSSSNIRLLQOQQLTTLTENSQSSPSVQWKGFGKSGHGQGNLSLSPLELDQSGTW 179
QY 183 TCTYLQNKQKVEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMQA 242
DB 180 TCIISQKTVKEFNIINLVAFQKVSMTFVARBEDQVEFSPLSPEDENLV--GELRQQA 237
QY 243 ERASSKSWITFEDLKNKESVYKRVTDQPKQMKKLPHLTPQALPOYAGSGNLTALAE 302
DB 238 QGASSSLIMISFILTENKLSKKAHAPLQMKESLPRLFTPLVYLSRASSGILTLML- 296
QY 303 AKTGKLRQEVNLVVMRATQLOKNTLCEWGPSPDKMLSLKENKEAKVSKREKPVWVNL 362
DB 297 AK-GTIVQEVNLVVMRANSSQNNLTCEVLTGPTSELTLSNLTKEQAQKVSQKQLVWVVD 355
QY 363 PEAGMQCLSDSGQVLTLESNIKVLPTWSTVPVPRASALPAPTGSALEPDP 422
DB 356 PEGGTWQCCLSDKDKVLAASLVN-----SSPV----- 383
QY 423 PASALPALA-----VISFLGLGLGVAC 447
DB 384 -VTSWPKFLAITLIGTILGLLIGLCVFC 412

RESULT 6
A27449
T-cell surface glycoprotein CD4 precursor - rat
NjAlternate names: W3/25 antigen
CjSpecies: Rattus norvegicus (Norway rat)
CjDate: 21-May-1988 #sequence_revision 21-May-1988 #text_change 21-Jan-2000
CjAccession: A27449; A35433
RjClark, S.J.; Jefferies, W.A.; Barclay, A.N.; Gagnon, J.; Williams, A.F.
Proc. Natl. Acad. Sci. U.S.A. 84, 1649-1653, 1987
AjTitle: Peptide and nucleotide sequences of rat CD4 (W3/25) antigen: evidence for deriv
AjReference number: A27449; MUID:87175535; PMID:3104900
AjAccession: A27449
AjMolecule type: mRNA
AjResidues: 1-457 <CLA>
AjCross-references: GB:M5768; NID:9203387; PIDN:AAA40901.1; PID:9203388
RjDavis, S.U.; Ward, H.A.; Puklaavec, M.J.; Willis, A.C.; Williams, A.F.; Barclay, A.N.
J. Biol. Chem. 265, 10410-10418, 1990
AjTitle: High level expression in Chinese hamster ovary cells of soluble forms of CD4 T
AjReference number: A35433; MUID:90285164; PMID:2113054
AjContents: annotation
CjSuperfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
CjKeywords: glycoprotein; membrane protein; surface antigen
Fj219-300/Domain: immunoglobulin homology <IMM>

Query Match      43.0%; Score 999; DB 2; Length 457;
Best Local Similarity 52.9%; Pred. No. 7.4e-59;
Matches 207; Conservative 61; Mismatches 115; Indels 8; Gaps 5;

QY 1 MNRGVPRHL-LIVLQALIPATQGNKVLGKGGDTVELTCTASQKSIQFHWKNSNQ 58
DB 1 MCRGFSFRHLPLLLQLSLKLVVQGTIVLQKGGASAEIPCESTSRSSAFAMKSSDQ 60
QY 59 IKILGNQGSFLTGKPSKLNDRADSRSLMDQGNPPLTIKUKIEDSDTYICEVDQKEV 118
DB 61 KTLIGYKNNKLTKISLELYSPDSRKAWERGSPLINKLRMDSDGTIVYCELENKEEV 120
QY 119 QLVFGITANSPTHLLOGQSITLTLESPPGSSPSVQCRSPRGKNIQSGKTLVSQLELQ 177
DB 121 ELWFRITPNRRLHLHGQSITLTLDSPKVSDDPICRKSNIIVDSKAFTHSLRIQ 180
QY 178 DSGTWTCTV-LQNKQVEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGE 237
DB 178 DSGTWTCTV-LQNKQVEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGE 237

```


Db 181 DSGINMCTVTLNOKKHSFDKMLSVLGFASTSTAYKSEGESAEFFSPFLNLGEESL--QGE 238
QY 238 LWMQARASSSSKSMITFDLKNKEVSVKRYVQDPKLGKGLPLHLTLPLQALPOYASSGNL 297
Db 239 LKRAKAEAPSSQSWITFSLTKNQKVSQKSTNPKFQSLFPLTLLOIPOVSLQFPASSGNL 298
QY 298 TLLEAKTKGLHDEVNLVVRAATQLQKN-LTCEVWGPTSPKMLSLKLENKAEKVSREK 356
Db 299 TLTLTD--RGLVGEVNLVVMKVTQPDSONLTICEVWPTSPKMLILKQENGEARVSROEK 356
QY 357 PVMVNLPEAGMWOCLLSDSGQVLLSENIKYL 387
Db 357 VIQVQAEAGVMQCLLSEGEVKKMDSKIQVL 387
RESULT 7
T-cell surface glycoprotein CD4 precursor - mouse
N:Alternate names: T-cell differentiation antigen L3T4; T-cell surface antigen T4/Len 3
C:Species: Mus musculus (house mouse)
C>Date: 30-Jun-1987 #sequence, revision 30-Jun-1987 #text, change 16-Jul-1999
C/Accession: A02110; A26038; A39893; A39955; 154564; 169018; A47642
R/Tourville11e, B.; Gorman, S.D.; Field, E.H.; Hunkapiller, T.; Parnes, J.R.
Science 234, 610-614, 1986
A>Title: Isolation and sequence of L3T4 complementary DNA clones: expression in T cells
A/Reference number: A02110; MUID:87018845; PMID:3094146
A/Accession: A02110
A/Molecule type: mRNA
A/Residues: 1-457 <TOU>
A/Cross-references: GB:M13816; NID:g192070; PIDN:AAA37267.1; PID:g309112
R/Litman, D.R.; Gettner, S.N.
Nature 325, 453-455, 1987
A>Title: Unusual intron in the immunoglobulin domain of the newly isolated murine CD4 (L3T4) gene
A/Reference number: A26038; MUID:87115821; PMID:3027575
A/Accession: A26038
A/Molecule type: mRNA
A/Residues: 1-457 <LIT>
A/Cross-references: GB:X04836; NID:g50353; PIDN:CAA8539.1; PID:g50354
R/Gorman, S.D.; Tourville11e, B.; Parnes, J.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 7644-7648, 1987
A>Title: Structure of the mouse gene encoding CD4 and an unusual transcript in brain.
A/Reference number: A39893; MUID:88041159; PMID:2823269
A/Accession: A39893
A/Molecule type: DNA
A/Residues: 1-25, 'E', 27-457 <GOR>
A/Cross-references: GB:M17080; GB:J03003; NID:g192515; PIDN:AAA37402.1; PID:g387124
R/Maddon, P.J.; Moliniaux, S.M.; Maddon, D.E.; Zimmerman, K.A.; Godfrey, M.; Alt, F.W.;
Proc. Natl. Acad. Sci. U.S.A. 84, 9155-9159, 1987
A>Title: Structure and expression of the human and mouse T4 genes.
A/Reference number: A39955; MUID:88097446; PMID:3501122
A/Accession: A39955
A/Status: nucleic acid sequence not shown; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 25-457 <MAD>
A/Note: the cited GenBank accession number, J03564, is not in release 101.0
R/Parnes, J.R.; Hunkapiller, T.
Immunol. Rev. 100, 109-127, 1987
A>Title: L3T4 and the immunoglobulin gene superfamily: New relationships between the immunoglobulin and the immunoglobulin gene superfamily.
A/Reference number: 154564; MUID:88152875; PMID:3326818
A/Accession: 154564
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-457 <RES>
A/Cross-references: GB:M36650; NID:g198670; PIDN:AAA39401.1; PID:g198671
A/Accession: 169018
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 208-318 <RE2>
A/Cross-references: GB:M36651; NID:g198672; PIDN:AAA39402.1; PID:g554183
R/Claesson, B.J.; Tsegataroglou, J.; Kitzbaum, L.; Maddox, J.; Mackay, C.R.; Brandon, M.; W
Immunogenetics 23, 129-132, 1986
A>Title: The L3T4 antigen in mouse and the sheep equivalent are immunoglobulin-like.
A/Reference number: A47642; MUID:8616694; PMID:3082751

A/Accession: A47642
A/Molecule type: protein
A/Residues: 27-43 <CLA>
C/Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells &
C/Genetics:
A/Introns: 18/1; 74/1; 128/1; 207/1; 319/1; 386/1; 425/3; 448/2
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
C/Keywords: alternative initiators; duplication; glycoprotein; T-cell; transmembrane pro
F1/26/Domain: signal sequence #status predicted <Sig>
F1/27-57/Domain: T-cell surface glycoprotein CD4 #status experimental <MAT>
F1/35-114/Domain: immunoglobulin homology <IM1>
F1/39-190/Domain: immunoglobulin homology <IM2>
F1/220-301/Domain: immunoglobulin homology <IM3>
F1/241-457/Product: CD4, brain-specific short form #status predicted <BRA>
F1/321-372/Domain: immunoglobulin homology <IM4>
F1/395-419/Domain: transmembrane #status predicted <TM>
F1/420-457/Domain: intracellular #status predicted <INT>
F1/42-112, 159-188, 328-370/Dissulfide bonds: #status predicted
F1/187, 296, 323, 392/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 42.7%; Score 993; DB 1; Length 457;
Best Local Similarity 53.8%; Pred. No. 1.9e-58;
Matches 211; Conservative 62; Mismatches 109; Indels 10; Gaps 7;
QY 1 MKRGVPRH-LTLVQLALPLPAATQGNKVLGKGGDTVELCTASOKKSIOFMKNSNOI 59
Db 1 MKRAISLRRLHLHLQLSLQSLAVTQKTLVLEKGESEAEPLPSSOKITFTWKFSDOR 60
QY 60 KILGNQGSFLTKG--PSKLNDRADSRRLMDQGNFPLIRKMLKIEDSPYICEVEDQKE 116
Db 61 KILGQHGKGVLRIGGSPSGF-DRPDSKKAMKSGSPULINKLKMEDSGTYICELNRKE 119
QY 117 EYQLVFGILTANSPTHLQGSGLTLTLES-PRGSSPSVQCKSPRKNIQGGTLYSQLE 175
Db 120 EVELWVFKVTFSFGTSLQGSGLTLTLDNSKVSNPLETECKRKQKGVSGSYLSKSNLR 179
QY 176 LDDSGTWTCTVQLQNKQVEFKIDIVLAFQKASIVYKKEGQVESPFLATVLELNGS 235
Db 180 VDDSDFMNCTVTLDDQKMFGLTSLVGLFQSAITRAYSEGSABEFSPFLNABE--NQM 237
QY 236 GELMWQARASSSSKSMITFDLKNKEVSVKRYVQDPKLGKGLPLHLTLPLQALPOYASSGN 295
Db 238 GELMWKAEKDSFQGWISFISIKNKEVSQKSTKDKLQKERTLPLTKIPOVSLQFPASSGN 297
QY 296 TLLEAKTKGLHDEVNLVVRAATQLQKNLTCEVWGPTSPKMLSLKLENKAEKVSREK 355
Db 298 TLTLTD--KGLHDEVNLVVWKAQNLNTLTCEVWGPTSPKMLILKQENGEARVSROEK 355
QY 356 KPMVNLPEAGMWOCLLSDSGQVLLSENIKYL 387
Db 356 KVVQVAPETGLMOCCLLSEGDVKVMDSRIOVL 387
RESULT 8
160082
CD4 receptor - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 29-May-1998 #sequence, revision 29-May-1998 #text, change 23-Jul-1999
C/Accession: 160082
R/Zverev, V.V.; Sidorov, A.V.; Nedospaev, S.A.; Malishova, V.V.; Udalova, I.A.; Andzha
Vopr. Virozol. 40, 100-102, 1995
A>Title: [Nucleotide sequence of two exons of the human T-lymphocyte CD4 receptor gene]
A/Reference number: 160082; MUID:95407135; PMID:767667
A/Accession: 160082
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-71 <RES>
A/Cross-references: GB:S79267; NID:g1086922; PIDN:AA835273.1; PID:g1086923
C/Genetics:
A/Introns: 17/1
C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
Query Match 15.4%; Score 357; DB 2; Length 71;

Best Local Similarity 98.6%; Pred. No. 2,5e-17;
Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NMRGVPFRHLVLVLTALPATQGNKVLGKKGDTVELTCASSQKKSIQFMKNSQIK 60

DB 1 NMRGVPFRHLVLVLTALPATQGNKVLGKKGDTVELTCASSQKKSIQFMKNSQIK 60

QY 61 ILNGSGFLTK 71

DB 61 ILNGSGFLTK 71

RESULT 9

S21461 T-cell surface glycoprotein CD4 (allele 1) - pig (fragment)

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 21-Jan-2000

C:Accession: I47131; S21461

R:Guatfesson, K.; Germana, S.; Sundt, T.M.

J. Immunol. 151, 1365-1370, 1993

A:Title: Extensive allelic polymorphism in the CDR2-like region of the miniature swine C

A:Reference number: I47131; MUID:93329116; PMID:833533

A:Accession: I47131

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-99 <GU2>

A:Cross-references: EMBL:X65629; NID:g1928; PIDN:CAA46583.1; PID:g388232

C:Superfamily: T-cell surface glycoprotein CD4; Immunoglobulin homology

C:Keywords: glycoprotein; T-cell

F/3-81/Domain: Immunoglobulin homology <IMM>

Query Match 13.1%; Score 305.5; DB 2; Length 99;

Best Local Similarity 60.2%; Pred. No. 1e-13;

Matches 59; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 32 KKGDTVELTCTASQKSIQFMKNSQIKILNGSGFL-TKGPSKLNDRADSRSLMDQ 90

DB 1 KAGDLAEIPCHSSQKKNLFFNMKNSQIKILGHSFMTASVTELSRLDSKKMWDHG 60

QY 91 NFPLIIKNIKEDSDTYICEVEDQKEVQLVFGTLAN 128

DB 61 SFPLIIKNIKEDSDTYICEVEDQKEVQLVFGTLAN 98

RESULT 10

A39016 T-cell surface glycoprotein CD7 precursor - human

M:Alternate names: T-cell leukemia antigen

C:Species: Homo sapiens (man)

C:Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 05-Nov-1999

C:Accession: A39016; S03520

R:Schenberg, L.E.; Flenor, D.E.; Kurtzberg, J.; Haynes, B.F.; Kaufman, R.E.

Proc. Natl. Acad. Sci. U.S.A. 88, 603-607, 1991

A:Title: Isolation and characterization of the genomic human CD7 gene: structural simila

A:Reference number: A39016; MUID:91110576; PMID:1703303

A:Accession: A39016

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-240 <SCH>

A:Cross-references: GB:M37271; NID:g180163; PIDN:AAA51953.1; PID:g180164

R:Artuffo, A.; Seed, B.

EMBO J. 6, 3113-3116, 1987

A:Title: Molecular cloning of two CD7 (T-cell leukemia antigen) cDNAs by a COS cell expr

A:Reference number: S03520; MUID:88111517; PMID:3501369

A:Accession: S03520

A:Molecule type: mRNA

A:Residues: 1-240 <ARU>

A:Cross-references: EMBL:X06180; NID:g29819; PIDN:CAA29546.1; PID:g732757

C:Genetics: GDB:CD7

A:Cross-references: GDB:119770; OMIM:166820

A:Map position: 17q25.2-17q25.3

A:Introns: 28/1

C:Keywords: glycoprotein; membrane protein; surface antigen; tandem repeat

F/1-25/Domain: signal sequence #status predicted <SIG>

F/26-240/Product: T-cell surface glycoprotein CD7 #status predicted <MAT>

F/145-180/Region: 9-residue repeats (X-P-P-X-A-S-A-L-P)

Query Match 12.6%; Score 293; DB 2; Length 240;

Best Local Similarity 40.3%; Pred. No. 2.2e-12;

Matches 91; Conservative 18; Mismatches 49; Indels 68; Gaps 9;

QY 270 PKLQMKKRLHLPLLPQALPYAGSNLTLALEKTKLHQEVNLVYMBARATQLKNLTCE 329

DB 5 PRLTL---LPLLLALAGLPg-----ALAAQEVQSGPHCTTPVGAS---VNITCS 49

QY 330 VMGPTSPKMLSLKLNKNAKVSRRKRPVWVLANPEAG----- 366

DB 50 TSGGLRGYLRQL-----GPGQDIIYVEDGVPTTDRFRGRIDFSSGQNLIT 98

QY 367 --WMQCLISDSG---QVLLESNI-----KVLPT-----W-STPVHPRASALPAPPTG 407

DB 99 ITMRLQLSDGTGTCAITEVNVYSGTLVLVTEQSGQWHRCSDAAPPASALPAPPTG 158

QY 408 SALPDQPTASALPDPASALPALAVISFLGLGACVLAARTR 453

DB 159 SALPDQPTASALPDPASALPALAVISFLGLGACVLAARTQ 204

RESULT 11

S21462 T-cell surface glycoprotein CD4 (allele 2) - pig (fragment)

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 20-Feb-1995 #sequence_revision 19-Apr-1996 #text_change 21-Jan-2000

C:Accession: I47132; S21462

R:Guatfesson, K.; Germana, S.; Sundt, T.M.

J. Immunol. 151, 1365-1370, 1993

A:Title: Extensive allelic polymorphism in the CDR2-like region of the miniature swine C

A:Reference number: I47131; MUID:93329116; PMID:833533

A:Accession: I47132

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-99 <GU2>

A:Cross-references: EMBL:X65630; NID:g1929; PIDN:CAA46584.1; PID:g388233

C:Superfamily: T-cell surface glycoprotein CD4; Immunoglobulin homology

C:Keywords: glycoprotein; T-cell

F/3-81/Domain: Immunoglobulin homology <IMM>

Query Match 12.1%; Score 280.5; DB 2; Length 99;

Best Local Similarity 56.1%; Pred. No. 4.6e-12;

Matches 55; Conservative 19; Mismatches 23; Indels 1; Gaps 1;

QY 32 KKGDTVELTCTASQKSIQFMKNSQIKILNGSGFLTKGP-SKLNDRADSRSLMDQ 90

DB 1 KAGDLAEIPCHSSQKKNLFFNMKNSQIKILGHSFMTASVTELSRLDSKKMWDHG 60

QY 91 NFPLIIKNIKEDSDTYICEVEDQKEVQLVFGTLAN 128

DB 61 SFPLIIKNIKEDSDTYICEVEDQKEVQLVFGTLAN 98

RESULT 12

JS0675 vascular cell adhesion molecule-1 precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 05-Nov-1999

C:Accession: JS0675; S19872; S23136

R:Hessom, C.; Moy, P.; Tizdat, R.; Chisholm, P.; Williams, C.; Wyse, M.; Burdly, L.; Mi

Biochem. Biophys. Res. Commun. 183, 163-169, 1992

A:Title: Cloning of murine and rat vascular cell adhesion molecule-1.

A:Reference number: JS0674; MUID:92181437; PMID:1371918

A:Accession: JS0675

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-739 <HES>

A:Cross-references: GB:M4488; NID:g207642; PIDN:AAA42332.1; PID:g207643

R:Williams, A.; Atkins, R.; Fries, J.; Gimbrone, M.A.; Cybulsky, M.I.; Collins, T. submitted to the EMBL Data Library, February 1992
 A>Description: Nucleotide sequence of rat vascular cell adhesion molecule-1.
 A:Reference number: S19872
 A:Accession: S19872
 A:Molecule type: mRNA
 A:Residues: 1-2, 'G', 4-121, 'HL', 124-165, 'N', 167-738, 'G' <MIL>
 A:Cross-references: EMBL:X63722; NID:957471; PIDN:CAA45254.1; PID:957472
 R:Williams, A.J.; Atkins, R.C.; Fries, J.W.U.; Gimbrone Jr., M.A.; Cybulsky, M.I.; Collins, T.
 A:Title: Nucleotide sequence of rat vascular cell adhesion molecule-1 cDNA.
 A:Reference number: S23136; MUID:92305064; PMID:1377031
 A:Accession: S23136
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2, 'G', 4-165, 'N', 167-738, 'G' <MIL>
 C:Comment: This protein interacts with the beta-1 integrin very late antigen 4 on leukocytes.
 C:Gene: VCAM-1
 C:Superfamily: immunoglobulin homology
 C:Keywords: cell adhesion; transmembrane protein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:15-739/Product: vascular cell adhesion molecule 1 #status predicted <VAS>
 F:239-293/Domain: immunoglobulin homology <IMM1>
 F:328-385/Domain: immunoglobulin homology <IMM2>
 F:526-581/Domain: immunoglobulin homology <IMM3>
 F:676-696/Domain: transmembrane #status predicted <TRA>
 F:697-715/Domain: intracellular #status predicted <INT>

Query Match 7.0%; Score 162; DB 2; Length 739;
 Best Local Similarity 22.0%; Pred. No. 0.005;
 Matches 86; Conservative 60; Mismatches 155; Indels 90; Gaps 14;

```

QY 33 KQDTVELTCTASQKSIQFHWNKSNQIKLNGCSFLTGPSKLNDRADSRSLMDQGNF 92
DB 238 EGAALVTMTASBGLPAPELFWKSK-----LDNGVQLD-----SGNA 274
QY 93 PLIKKLIKEDSDTYICE-----VEDQKEVOVLV-----FGLTNSDTHLLQGOSLTLL 142
DB 275 TLTLLMRHEDSGIYCEGVNLVGRKTEVELVQKRPPTVDDSPSGVAAQVDSVLT 334
QY 143 LSPSPSSPSVQCRSPRGKNIQ-----GKTLVSQLELDQSGTWTCTVLQNKKEVF 195
DB 335 CAAVGDSDFSPSRMTGTDSPLNCEVADGATSTLTLSPPGVDEHSHYLTCTVQGRKLEK 394
QY 196 KXDIVVLAQKASSIYKKEGGEVERSPFLA-----FTVEKLTGSG 236
DB 395 TLOVEYVSF-----PEDPEIEISGPLVHGRPTVAVCPVNYPPDHLLEIELKGET 445
QY 237 ELMMQKERS-SSKSWITFDLKNKEYSVKRVQDPRLQNGKLT-----PLHLTLPQALP-Q 290
DB 446 TLINKLREIEIGKIS--LETKSLMTPTPTAD---TGKALVCLAKLHSSQMESEPPQ 498
QY 291 VAGSGNLTALAEKTKGLKQEVNLVYMRATOLQKNTCEWGPSTSPKMLSKLEKKEAK 350
DB 499 RQSTQTLVYVNAPEKRTIIVSPSPVPEBSGPV--NITCSDEPPTKIIMSNQKNGELQ 556
QY 351 VSKREKPVVNLPEAGMOCCLSDSQVILE 381
DB 557 PLSQ-----NTTSLFMATKEDSGIYCE 580

```

RESULT 13
 JJC2457
 vascular cell adhesion protein - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 15-Feb-1995 #sequence_revision 05-Apr-1995 #text_change 08-Oct-1999
 C:Accession: JJC2457
 R:Tsang, Y.T.M.; Haskard, D.O.; Robinson, M.K.
 Biochem. Biophys. Res. Commun. 201, 805-812, 1994
 A:Title: Cloning and expression kinetics of porcine vascular cell adhesion molecule.
 A:Reference number: JJC2457; MUID:94271236; PMID:7516159
 A:Accession: JJC2457

A:Molecule type: mRNA
 A:Residues: 1-538 <TSA>
 A:Cross-references: EMBL:U08351; NID:g474382; PIDN:AAA21542.1; PID:g474383
 C:Keywords: glycoprotein; transmembrane protein
 F:497-517/Domain: transmembrane #status predicted <TM>
 F:75,157,271,330,360/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.7%; Score 155.5; DB 2; Length 538;
 Best Local Similarity 23.4%; Pred. No. 0.0088; Mismatches 91; Indels 73; Gaps 11;
 Matches 65; Conservative 49;

```

QY 32 KKQDTVELTCTASQKSIQFHWNKSNQIKLNGCSFLTGPSKLNDRADSRSLMDQGN 91
DB 235 QEGDSMMWMTCTSEGLPAPQISW-----SKLDNDQDQL-----SGN 271
QY 92 PLIKKLIKEDSDTYICE-----VEDQKEVOVLV-----FGLTNSDTHLLQGOSLTLL 141
DB 272 ATLTLIAMEDESGIYCEGVNVPVGTNRKEVELTQVAPRDTTISVNPSTLEGSVVM 331
QY 142 TLESPGSSPSV-----QCRSPRGKNIQSGKTLVSQLELDQSGTWTCTVLTQ-----NOKV 193
DB 332 TCGSGFPAPKILMSKRLDGNLEPLSENVTLLTSTYKEDSGIYCEGINQAGINRKEV 391
QY 194 EFKI-----DIVVLAQKAS-----SIYKKE-----GEQVERSPFLA 226
DB 392 ELILQAPKDQLTLPSPSSVYEGPTVISCCTCGVNPRTLILKKAETGDTVLKSTDA 451
QY 227 FTVEKL---TSGELMMQKERSASSKSWITFDLNKE 260
DB 452 YTHRAPLADAGVCESSKNEIGQLRS-ITLDVGR 488

```

RESULT 14
 AA1288
 vascular cell adhesion molecule 1, long splice form precursor - human
 N:Alternate names: VCAM-1
 C:Species: Homo sapiens (man)
 C:Date: 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 31-Jan-2000
 C:Accession: AA1288; S11476; A39755; B39755; A61160; A43352; F81379; A39554
 R:Cybulsky, M.I.; Fries, J.W.U.; Williams, A.J.; Sultan, P.; Eddy, R.; Byers, M.; Shows, Proc. Natl. Acad. Sci. U.S.A. 88, 7859-7863, 1991
 A:Title: Gene structure, chromosomal location, and basis for alternative mRNA splicing
 A:Reference number: AA1288; MUID:91352090; PMID:1715583
 A:Accession: AA1288
 A:Molecule type: DNA
 A:Residues: 1-739 <CYB>
 A:Cross-references: GB:M73255; NID:g340195; PIDN:AAA61270.1; PID:g340196
 R:Polte, T.; Newman, W.; Gopal, T.V.
 Nucleic Acids Res. 18, 5901-1990
 A:Title: Full length vascular cell adhesion molecule 1 (VCAM-1).
 A:Reference number: S11476; MUID:91016951; PMID:1699207
 A:Accession: S11476
 A:Status: translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-739 <POU>
 A:Cross-references: EMBL:X53051; NID:g37648; PIDN:CAA37218.1; PID:g37649
 R:Hession, C.; Tizard, R.; Vassallo, C.; Schiffer, S.B.; Goff, D.; Moy, P.; Chi-Rosso, G
 J. Biol. Chem. 266, 6682-6685, 1991
 A:Title: Cloning of an alternate form of vascular cell adhesion molecule-1 (VCAM1).
 A:Reference number: A39755; MUID:91201302; PMID:1707873
 A:Accession: A39755
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 25-646, 648-739 <HEB1>
 A:Cross-references: GB:M60335
 A:Note: the complete translation is not shown
 A:Accession: B39755
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-739 <HEB2>
 A:Cross-references: GB:M60335; NID:g340193; PIDN:AAA61269.1; PID:g340194
 A:Experimental source: cell type endothelial cell; tissue type umbilical vein; map 1032-R:Cybulsky, M.I.; Fries, J.W.U.; Williams, A.J.; Sultan, P.; Davis, V.M.; Gimbrone Jr.,

Am. J. Pathol. 138, 815-820, 1991
 A>Title: Rapid communication. Alternative splicing of human VCAM-1 in activated vascular
 A:Reference number: A61160; MUID:91189297; PMID:1707234
 A:Accession: A61160
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 25-401; 'T', 403-686 <CY2>
 R:Ademaco, M.F.; McQuillan, J.J.; Rosen, G.D.; Dean, D.C.
 J. Biol. Chem. 267, 16323-16329, 1992
 A>Title: Characterization of the promoter for vascular cell adhesion molecule-1 (VCAM-1)
 A:Reference number: A43352; MUID:92355594; PMID:1379595
 A:Accession: A43352
 A:Molecule type: DNA
 A:Residues: 1-21 <IAD>
 A:Cross-references: GB:M92431; NID:9340197
 A>Note: sequence extracted from NCBI backbone (NCBI:110680, NCBI:110681)
 R:Obborn, L.; Vassallo, C.; Benjamin, C.D.
 J. Exp. Med. 176, 99-107, 1992
 A>Title: Activated endothelium binds lymphocytes through a novel binding site in the al
 A:Reference number: PH1379; MUID:92308860; PMID:1377228
 A:Accession: PH1379
 A:Molecule type: protein
 A:Residues: 25-181; 'G', 183-402 <OSB>
 C:Comment: This adhesion molecule is induced on endothelial cells by inflammatory cyto
 C:Genetics:
 A:Gene: GDB:VCAM1
 A:Cross-references: GDB:127922; OMIM:192225
 A:Map position: 1p32-1p31
 C:Keywords: alternative splicing; cell adhesion; glycoprotein; transmembrane protein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-739/Product: vascular cell adhesion molecule 1, long splice form #status predicted
 F:25-698/Domain: extracellular #status predicted <EXT>
 F:699-720/Domain: transmembrane #status predicted <TM>
 F:721-739/Domain: intracellular #status predicted <INT>
 F:273,365,417,463,531,561/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.7%; Score 155; DB 2; Length 739;
 Best Local Similarity 21.9%; Pred. No. 0.015;
 Matches 91; Conservative 71; Mismatches 166; Indels 80; Gaps 18;

QY	8	RLHLVQLALPAAT--OGNKNVLGKKDYTELTCTASQKSIQPHKNSQITLGNQ	65
DB	211	RQAVKELQVYISPKKTVISVNPSTKLQEGSVTMTCSSEGLPAPRIFW-----	258
QY	66	GSFLTKPSKLNDRSRSLMDQGNPLIINKLIEDDTYCE-----VEDQKEVQL	121
DB	259	-----SKKLDNGLQLH---SGNATTLTAMKEDSGIYCEGVNLIGKRRKEVELI	307
QY	122	V-----FGLTANSPTHLLQ--GQSLTTLTSSPPGSSPVQCRSPRKNIOG-----GKT	168
DB	308	VOEKPTVAISPGPIAAGIDGSMVLTGVCESPSFMRQIDSPISGVKRSQSTNST	367
QY	169	LSVSQLELDSDGTWCTCTVQLQNKVYFKDIYVLAQKASSIVYK---KEGQVEFS---	222
DB	368	LTLSPVSEFNEHSYCTVYCGHKLEKGIQVELYFPRPDELMSGVLNGSSVTSKV	427
QY	223	---PPL-AFTVVKLTGSGELMWQABERASSKSWTF---DLKNKVSVKRTQDPKQ	273
DB	428	BSVYPLDRLEILKGETIL-----ENIEFLETDKKSLEKNSLEMTFLPTIED	476
QY	274	MGKCL---PLHLTLPQALPOVAGSGNTLALAEATGKLHDEVLNVKRAQLOK---N	325
DB	477	TKKALVCAQAKHIDMEFEPRKQRO---TQTLVYVAVP--RDTYLVSPSSILBEGSSVN	531
QY	326	LTCBWGPTSPKLMLSKLEKAKVSKREKPVWLNPPAGMWQCLSDSGVLLF	381
DB	532	MTCLSQGFPAKILMSRQLPENGLOPLSENATLTISTK-----MEDSGVLYCE	580

RESULT 15
 TCHUNG
 neural cell adhesion molecule 1 GPI-anchored splice form precursor, muscle-specific - hu
 N:Alternate names: CD56; NCAM-120

C:Species: Homo sapiens (man)
 C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 28-Jan-2000
 C:Accession: S07784; #sequence
 R:Barton, C.H.; Dickson, G.; Gower, H.J.; Rowett, L.H.; Pult, W.; Elsom, V.; Moore, S.E.
 Development 104, 165-173, 1988
 A>Title: Complete sequence and in vitro expression of a tissue-specific phosphatidylinos
 A:Reference number: S07784; MUID:93305258; PMID:3253057
 A:Accession: S07784
 A:Molecule type: mRNA
 A:Residues: 1-761 <BAR>
 A:Cross-references: EMBL:X16841; NID:935005; PIRN:CA34739.1; PID:935006
 R:Dickson, G.; Gower, H.J.; Barton, C.H.; Prentice, H.M.; Elsom, V.L.; Moore, S.E.; Cox,
 Cell 50, 1119-1130, 1987
 A>Title: Human muscle neural cell adhesion molecule (N-CAM): identification of a muscle-
 A:Reference number: A90895; MUID:87301755; PMID:2887295
 A:Accession: A26883
 A:Molecule type: mRNA
 A:Residues: 491-761 <DIC>
 A:Cross-references: GB:M17409; NID:9189097; PIDN:AAA5912.1; PID:9386979
 C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
 C:Genetics:
 A:Gene: GDB:NCAM1; NCAM; CD56
 A:Cross-references: GDB:119448; OMIM:116930
 A:Map position: 11q22.2-11q22.3
 C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
 C:Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membrane
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-761/Product: neural cell adhesion molecule phosphatidylinositol-linked form, muscle
 F:34-98/Domain: immunoglobulin homology <IMM1>
 F:132-191/Domain: immunoglobulin homology <IMM2>
 F:152-166/Region: heparin binding #status predicted
 F:161-165/Region: heparin binding #status predicted
 F:228-289/Domain: immunoglobulin homology <IMM3>
 F:263-272/Region: NCAM binding #status predicted
 F:322-387/Domain: immunoglobulin homology <IMM4>
 F:419-481/Domain: immunoglobulin homology <IMM5>
 F:499-587/Domain: fibronectin type III repeat homology #status atypical <FN3A>
 F:633-720/Domain: fibronectin type III repeat homology #status atypical <FN3B>
 F:41-96, 139-189, 235-287, 329-385, 426-479/diulfide bonds: #status predicted
 F:222,315,347,423,449,478/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.2%; Score 145; DB 1; Length 761;
 Best Local Similarity 20.5%; Pred. No. 0.07;
 Matches 100; Conservative 67; Mismatches 211; Indels 110; Gaps 20;

QY	3	RG-VPRHLLVLOLALPAATQGNKVLGKKDYTELTCTASQKSIQPHKNSQITKI	61
DB	196	RGEINFDIOVTVVPPITQARQNI VATAATLQGSVTLVCDAGPPEPTSMW-----	247
QY	62	LGNGSFLTKPSKLNDRSRSLMDQGNPLIINKLIEDSDTYICEVE-----DQKEE	117
DB	248	-----TKDGEIQEEDDEKTIFFSDSSQLTIKKYDKDAEYICIAENKAGEQDAT	299
QY	118	VQLLVFG---LTRANSPTHLLQGSGLTTLTSSPPGSSPVQCR-----SPRKNIOG-	166
DB	300	IHLKVPKPKITTYENQTAHLEBQVTLTCEASDPIPSITMRSTNNISSEKTLQDM	359
QY	167	-----KTLVSQLELDSDGTWCTCTVLO--NOKVYFKDIYVLAQKASSIVYKKEG	216
DB	360	VVRSHARVSLTLKLSIYTTAGETICTASNTIGDSQSMYLEVOYAPKLOQPVAVYTWEG	419
QY	217	EQVE-----FSPPLAFTVVKLTGSGELMWQABERASSKSWTFPDKN-KEVSVKRVTD	270
DB	420	NOVAITEVFAYPSA-TIS-----WFRDQGLLPSNNYSNITKYVTPASYLEVTPDS	470
QY	271	KLQNGK-----KLPLHLTLPQALPOVAGSGNTLALAEATGKLHQ	310
DB	471	ENDFGNNCTAVNRIGQESLEFLVQADTPSSPSIDQVEPVSSTAQVQFDEPBAATG-----	526
QY	311	EVNLVVRATOLQNLTCBWGPTSPKLMLSKLEKAKVSKREKPVW--LNPPAG	366
DB	527	--GVPILKYKAEVAVGEVW-----HSKRVYDAKESMEGIYIVIGLKPEPTT	571

QY 367 MNCGLSDSGOVL-----SNIKVPTWSTPVHPASALPAPPTGASLPDPDQTSLAPDP 422
 Db 572 YAVRLAALNKGIGELISAASEFKTQPVHSP--PASASSTPVPLSPD---TWPLP 625
 QY 423 PAASALPA 430
 Db 626 ALATTEPA 633

RESULT 16

```
vascular cell adhesion molecule-1 long splice form precursor - mouse
C|Species: Mus musculus (house mouse)
C|Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
A|Accession: JN0581; JS0674; A40274; A48919
R|Araki, M.; Araki, K.; Vassalli, P.
Gene 126, 261-264, 1993
A|Title: Cloning and sequencing of mouse VCAM-1 cDNA.
A|Reference number: JN0581; MUID:93246254; PMID:7683304
A|Accession: JN0581
A|Molecule type: mRNA
A|Residues: 1-739 <ARA>
A|Cross-references: EMBL:X67783; NID:g298116; PIDN:CMA47989.1; PID:g298117
R|Hession, C.; Moy, P.; Tizard, R.; Chesholm, P.; Williams, C.; Wyse, M.; Burkle, L.; Mc
Biochem. Biophys. Res. Commun. 183, 163-169, 1992
A|Title: Cloning of murine and rat vascular cell adhesion molecule-1.
A|Reference number: JS0674; MUID:92181437; PMID:1371918
A|Accession: JS0674
A>Status: nucleic acid sequence not shown
A|Molecule type: mRNA
A|Residues: 1-739 <HES>
A|Cross-references: GB:M64487; NID:g202345; PIDN:AAA40545.1; PID:g202346
R|Miyake, K.; Medina, K.; Ishihara, K.; Kimoto, M.; Auerbach, R.; Kincaide, P.W.
J. Cell Biol. 114, 557-565, 1991
A|Title: A VCAM-like adhesion molecule on murine bone marrow stromal cells mediates bind
A|Reference number: A40275; MUID:91317872; PMID:1713592
A|Accession: A40275
A|Molecule type: protein
A|Residues: 'XX',27-32 <MIY>
R|Cybulsky, M.I.; Allam-Motamed, M.; Collins, T.
Genomics 18, 387-391, 1993
A|Title: Structure of the murine VCAM1 gene.
A|Reference number: A48919; MUID:94117008; PMID:7507076
A|Accession: A48919
A|Molecule type: DNA
A|Residues: 1-692, 'N',694-739 <CYB>
A|Cross-references: GB:I22355; NID:g347981; PIDN:AAI6921.1; PID:g459893; GB:I22301
C|Comment: This protein is a transmembrane protein and interacts with the beta-1 integrin
C|Superfamily: immunoglobulin homology
C|Keywords: alternative splicing; cell adhesion; glycoprotein; transmembrane protein
F|1-74/Domain: signal sequence #status predicted <SIG>
F|25-739/Product: vascular cell adhesion protein-1 #status predicted <MAT>
P|239-293/Domain: immunoglobulin homology <IMM1>
F|328-385/Domain: immunoglobulin homology <IMM2>
F|521-581/Domain: immunoglobulin homology <IMM3>
F|699-720/Domain: transmembrane #status predicted <TM>
F|225,273,424,531,Binding site: carbohydrate (Asn) (covalent) #status predicted
```

```

Db 331 VUTCAICGDSPPSSMRTQDSDPLANGVRNRGAKSTLVLSVGFEDBHSYLCAVTCIOR 390

Qy 192 KXEFKIDIVLAFORASSIVYKKEGEYFSPFLA-----FTVEKX 232
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 TLEKRTQVEVVSF-----PEDPVXMGSPVLRHAPVTNCTVBNPVYPFHLEIEL 441
   : : : : : : : : : : : : : : : : : : : : : : : :

Qy 233 TSGGELMWQAEBAASSKSMI-FTDLKXNEVSVKRVQDDPKIQ-MGKUL-----PLHLLTPQ 286
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 442 KG-----ETTLMKKTFLEBMGKLSLEKILETTFTPIETIEDGKSLVCLARHSGEME 493
   : : : : : : : : : : : : : : : : : : : : : : : :

Qy 287 ALFOYAGSNLTALAEATYGLKHQEVNLYVVRATOLQ-----NLTCEVMGPTSPKMLTSL 342
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 494 SEPKQNSQVPLVNVAP-----KETTIVWSPSPILBEGSPVNVLTGSSDGI PAKILMSR 548
   : : : : : : : : : : : : : : : : : : : : : : : :

Qy 343 KLENKEAKVSKREKVVWVLN--PEAGMOC 370
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 QLNNGELQPLSENTTLTJTEMSTKRDSGIIYVC 579
   : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 17

A:Vascular cell adhesion molecule 1, short splice form precursor - human
 N/Alternate names: VCAM-1
 C/Species: Homo sapiens (man)
 C/Date: 03-Apr-1992 #sequence_revision 14-Jul-1994 #text_change 31-Jan-2000
 C/Accession: B41288; A33758
 R/Cybulsky, M.I.; Fries, J.W.U.; Williams, A.J.; Sultan, P.; Eddy, R.; Byers, M.; Shows, Proc. Natl. Acad. Sci. U.S.A. 86, 7859-7863, 1991
 A>Title: Gene structure, chromosomal location, and basis for alternative mRNA splicing c
 A/Reference number: A41288; MUID:91352090; PMID:1715583
 A/Accession: B41288
 A:Molecule type: DNA
 A:Residues: 1-647 <CTB>
 A:Cross-references: GB:W73255
 R/Osborn, L.; Hession, C.; Tizard, R.; Vassallo, C.; Luhnshkyj, S.; Chi-Rosso, G.; Lobb, Cell 59, 1203-1211, 1989
 A>Title: Direct expression cloning of vascular cell adhesion molecule 1, a cytokine-indu
 A/Reference number: A33758; MUID:90090619; PMID:268888
 A/Accession: A33758
 A:Molecule type: mRNA
 A:Residues: 1-647 <OSA>
 A:Cross-references: GB:M30257; NID:g179885; PIDN:AAA51917.1; PID:g179886
 C/Keywords: alternative splicing; cell adhesion; glycoprotein; transmembrane protein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-64/Product: vascular cell adhesion molecule 1, short form #status predicted <INT>
 F:25-606/Domain: extracellular #status predicted <EXT>
 F:607-628/Domain: transmembrane #status predicted <TM>
 F:629-647/Domain: intracellular #status predicted <INT>
 F:273,325,371,439,469/Binding site: carbohydrate (Asn) (covalent) #status predicted

A:Reference number: 159098; MUID:87204247; PMID:3033671
A:Accession: 159098
A>Status: translated from GB/EMBL/DBSJ
A:Molecule type: mRNA
A:Residues: 331-702 <RE2>
A:Cross-references: GB:M16234; NID:q180240; PIDN:AA51972.1; PID:q180241
R:Stiepen, D.; Paxton, R.J.; Neumaier, M.; Shively, J.E.; Wagener, C.
Biochem. Biophys. Res. Commun. 147, 212-218, 1987
A:Title: Carcinoembryonic antigen (CEA) and two crossreacting antigens of 165 KD and 105
A:Reference number: A26831; MUID:87326349; PMID:3632664
A:Accession: A26831
A:Molecule type: protein
A:Residues: 35-64 <SIE>
R:Thomas, P.; Toth, C.A.
Biochem. Biophys. Res. Commun. 170, 391-396, 1990
A:Title: Carcinoembryonic antigen binding to Kupffer cells is via a peptide located at C
A:Reference number: A35490; MUID:90321257; PMID:2372297
A:Accession: A35490
A:Molecule type: protein
A:Residues: 'X', 140-151, 'X', 155-156 <THO>
A>Note: this is the amino terminal end of a fragment shown to mediate uptake by Kupffer
C:Comment: This heavily glycosylated membrane protein of unknown function is a widely use
C:Comment: This protein may be processed at its C-terminus. It is anchored to the membra
C:Genetics:
A:Gene: GDB:CEA
A:Cross-references: GDB:119054; OMIM:114890
A:Map position: 19q13.2-19q13.2
A:Features: 22/1, 142/1, 235/1, 320/1, 413/1, 498/1, 591/1, 676/1
C:Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi
C:Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphati
F:1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F:1-34/Domain: signal sequence #status predicted <SIS>
F:35-678/Product: carcinoembryonic antigen #status predicted <MAT>
F:160-211/Domain: immunoglobulin homology <IMM2>
F:252-301/Domain: immunoglobulin homology <IMM2>
F:338-395/Domain: immunoglobulin homology <IMM3>
F:516-573/Domain: immunoglobulin homology <IMM3>
F:608-657/Domain: immunoglobulin homology <IMM5>
F:679-702/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F:678/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Gly) (in mature form

Query Match 5.9%; Score 137.5; DB 2; Length 702;
Best Local Similarity 18.8%; Pred. No. 0.2;
Matches 96; Conservative 76; Mismatches 215; Indels 123; Gaps 20;

QY 20 PATQGNKVLGKGDVLELTCTASQKSIQFMKNSNOIKLGNQSFLLTGPSTKXLR 79
DB 147 PSLSSNNSKPVEDK-DAVAFTCEPFTQDATYLMWVNNQSLPV----- 187
QY 80 ADSRRLMOGNPPLIKLKIEDSDTYICEVD-----QKEVQL-LVFGILTANS---- 129
DB 188 --SPRLQLSNGKNTLLFNVTRNDTASYKCETONPVARSRSVILNVLVYGPDPAPISPL 245
QY 130 DTHLLQSGSLTLTLESPPSSPSPVQCR-SPRGKNIQGGKTLVSQLELQDSGMTCTVQ 188
DB 246 NTSYRGEMLNLSCHA--ASNPPAQSFWVNGTFQOSTOELPFPNTVNNSSSYTCQAHN 303
QY 189 NOKKV-EFKIDIVLAFQKASSIVYKKEGOVEFSPPLAFVE-KLTSGELMGAERAS 246
DB 304 SDPLGRLRTVTITVVAEPPKPFITSNNSNPVEDEDAVALTCEPELQNTTYLMWVNNQSL 363
QY 247 SSKSWITFLPKKKEVSKVKTQDP-----KLQMGKLLPLH-----TLPPA 287
DB 364 PVSPPRLQLSNDNTLLSVTRNDVGRYECGIGNELSVHSDPVLNVLVYGPDPAPISPS 423
QY 288 LPOYASGNLTLALAKT-----GKLHQEV-NLVVMRATQLOKNL-TCEWVGPT 334
DB 424 YTYTRGCVNLISCHAAANPPAQSFWVNGTFQOSTOELPFPNTVNNSSSYTCQAHN 483
QY 335 SPRLMLSLKLENKAVY-----SKREKPV-----WVLNPPAGMQLLSDSGQVLE 381
DB 484 SGHSRTTVVXTITVASLPLKPSISNNSKPVEDKDAVAFTCEPELQNTTYLMWVNNQ----- 539

QY 382 SNIKVLPTWSTVHPRASALPAPPT-----GSALPDPOTASAL 419
DB 540 -----SLPVSRLQLSNGKNTLLFNVTRNDARAVYCGIQNSVSNRSDPVTLDVL 590
QY 420 --PDPPAASALPALAVISFLIGLGLGVAC 447
DB 591 YGPDPIPIISPPDS-----SYLSGANILNISC 615

RESULT 20
S36126
neural cell adhesion molecule L1 - rat
N:Alternate names: nerve growth factor-inducible large external glycoprotein; NILE glyco
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 20-Aug-1999
C:Accession: S36126; S17655; A60917; A30326
R:Minura, M.; Kobayashi, M.; Asou, H.; Uyemura, K.
FEBS Lett. 289, 91-95, 1991
A:Title: Molecular cloning of cDNA encoding the rat neural cell adhesion molecule L1. Tw
A:Reference number: S17655; MUID:91372414; PMID:1894011
A:Accession: S36126
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1259 <MTU>
A:Cross-references: EMBL:X59149
A:Accession: S17655
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1178, 1183-1259 <MT2>
A:Cross-references: EMBL:X59149; NID:q56740; PIDN:CAA41860.1; PID:q56741
R:Prince, J.T.; Milona, N.; Stallcup, W.B.
J. Neurosci. 9, 1825-1834, 1989
A:Title: Characterization of a partial cDNA clone for the NILE glycoprotein and identifi
A:Reference number: A60917; MUID:89257627; PMID:2723751
A:Accession: A60917
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1159-1199, 'G', 1201-1235, 'K', 1237 <PRI>
A>Note: this paper appeared earlier, with printing errors, as reference A30326
R:Prince, J.T.; Milona, N.; Stallcup, W.B.
J. Neurosci. 9, 876-883, 1989
A:Title: Characterization of a partial cDNA clone for the NILE glycoprotein and identifi
A:Reference number: A30326; MUID:89177485; PMID:2466966
A:Contents: annotation
A>Note: This paper was reprinted as reference A60917 to correct the omission of several
C:Comment: This sequence of this surface-accessible glycoprotein differs at only two pos
C:Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homology; n
C:Keywords: cell adhesion; duplication; glycoprotein; membrane protein
F:531-592/Domain: immunoglobulin homology <IMM>

Query Match 5.8%; Score 135; DB 2; Length 1259;
Best Local Similarity 20.2%; Pred. No. 0.63;
Matches 66; Conservative 58; Mismatches 114; Indels 88; Gaps 13;

QY 10 LILVQLALLPATQGNKVLGKGDVLELTCTASQKSIQ--FWKNSNOIKLGNQGS 67
DB 507 ILANLQVBEARQITQGPSSTIEKKGARVTFQASFPDSLQASITWRDGR----- 557
QY 68 FLTGPSTKLNDRADRSRLMOGNPPLIKLKIEDSDTYIC---EVEDKEEVQLLVF 123
DB 558 -----DLQEKGDSDKFTIEDQ--LVTKSLDSYDQDYSQVASTEDLDEVSRAQLLV 608
QY 124 GLTAN-----SDTHLQSGSLTLTLESPPSSPSPVQCRSP-----RGKNIQGGKTL 169
DB 609 GSPGPVPHLESDRLRLKQSGVHLSW-----SPADHNSPIEKVDIEFDKEMAPKMF 662
QY 170 SYSQLELDQSGTWTTVYQNGKVEFKIDIVL-----APQKASI 210
DB 663 SLGKV-----PGNQSTTLKSLPYVATPRVVAINKYRGEDSPVSETVVTPEAAPEKPV 718
QY 211 VYKKEGEO-----VEFSFP-LAFTVE-KLTSGELMGAERAS-----SSSK 249

Db 719 DVGSGNETNNVITWKLRLWMDNAPQIQXRVQRPLGQETWKEQTVSDPFLVNSNTS 778
QY 250 SWTFEDLNKKEYSVKRVTPQDPKLQNG 275
Db 779 TFVPEIKVQAVNNOCKGPEFOVTTG 804

RESULT 21
S05479
neural cell adhesion molecule L1 precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 10-Sep-1999
C/Accession: S05479; B60850; S22167
R/Moon, M.; Tacke, R.; Scherer, H.; Teplow, D.; Frueh, K.; Schachner, M.
Nature 334, 701-703, 1988
A/Title: Neural adhesion molecule L1 as a member of the immunoglobulin superfamily with
A/Reference number: S05479; MUID:88318924; PMID:3412448
A/Accession: S05479
A/Molecule type: mRNA
A/Residues: 1-1260 <MOO>
A/Cross-references: EMBL:X12875; NID:953336; PID:CAA33368.1; PID:953337
A/Note: the authors translated the codon CCT for residue 166 as leu, ACT for residue 396
A/Note: part of this sequence, including the amino end of the mature protein, was confir
R/Rathjen, F.G.; Wolff, J.M.; Frank, R.; Bonhoeffer, F.; Rutishauser, U.
J. Cell Biol. 104, 343-353, 1987
A/Title: Membrane glycoproteins involved in neurite fasciculation.
A/Reference number: A60850; MUID:87109457; PMID:3805123
A/Accession: B60850
A/Molecule type: protein
A/Residues: 20-28, 'XX', '31-36 <RAT>
R/Kohl, A.; Giese, K.P.; Mohjerl, M.H.; Montag, D.; Moon, M.; Schachner, M.
submitted to the EMBL Data Library, December 1991
A/Description: Analysis of promoter activity and 5' genomic structure of the neural cell
A/Reference number: S22167
A/Accession: S22167
A/Molecule type: DNA
A/Residues: 1-165, 'L', 167-189, 'E', 191-281, 'S', 283-395, 'S', 397-514, 'ABEKNPVDV', 524, 'GEGNE
A/Cross-references: EMBL:X63511
C/Genetics:
A/Introns: 26/1, 31/1, 66/2, 133/1, 174/1, 231/1, 268/2, 330/1, 374/1, 422/1, 459/2
A/Note: the list of introns may be incomplete
C/Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homology;
C/Keywords: alternative splicing; cell adhesion; duplication; glycoprotein; transmembran
F.1-19/Domains: signal sequence #status predicted <SIG>
F.20-1260/Product: neural cell adhesion molecule #status experimental <MAT>
F.256-313/Domains: immunoglobulin homology <IMM1>
F.440-498/Domains: immunoglobulin homology <IMM2>
F.531-592/Domains: immunoglobulin homology <IMM3>

Query Match 5.8%; Score 134.5; DB 1; Length 1260;
Best Local Similarity 20.1%; Pred. No. 0.69;
Matches 73; Conservative 59; Mismatches 140; Indels 91; Gaps 15;

QY 34 GDTVELTCTAQSOKSIQFHWKNSNOIKILNGSFLTKGPSKLNDRADRSRLMDQGNP 93
Db 346 GETALDLCQVQRPQPEITWR-----INGMSMETVKNQDKYRI-BQGS-- 387
QY 94 LIIKLIKIEDSDTYICEVDOK-----EENGLVPGLTANSDTHL-DQGSLVTLTL 143
Db 388 LILSNVQPTDVTWTCCEARNQHGLLIANAYIVVLPARILTKDQVTMAVAGSTAYILLC 447
QY 144 ESPSSPSVQCRSPRGKNI-----QGGKTLVSQLELDGSGTWTCTVQLNQKVEF 195
Db 448 KAFGARPVESVQWLDEGTVLQDERFPFPAANGTILSIRDLQANDTGRYCOQANDQNTI 507
QY 196 KIDIVLAFQKASSI-----VYKKEGEVSEFPLAFTVEKLTSGSELMMQAEARASSK 249
Db 508 LANLVQ---KEATQITQGPBSAIEKKGAIVTFTQASFPSL---QASITWGDGR----- 557
QY 250 SWITFDLNKKEYSVKRVTPQDPKLQNGKULPLHLTLPQALPQVAGSGLNTLAEATGKLH 309
Db 558 -----DLQERGSDDKRYFIEDGLVTI-----QSL-DYSDQGNYSVASTELDEVE 600

QY 310 QEVLVYVWRA-----TQLQKULTCEVWGP-----TSPKLMLSLKLENKEAK 350
Db 601 SRAQLLVGSPGVPVPHLELSDRHLTKQSQVHLS---WSPADHNSPIEKYDIEFEDKEMA 657
QY 351 VSK 353
Db 658 PEK 660

RESULT 22
T34416
hypothetical protein F12F3.2 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence revision 29-Oct-1999 #text_change 29-Oct-1999
C/Accession: T34416
R/Fulton, B.; Mohlmann, P.
submitted to the EMBL Data Library, July 1998
A/Description: The sequence of C. elegans cosmid F12F3.
A/Reference number: 221521
A/Accession: T34416
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2783 <FUL>
A/Cross-references: EMBL:U80022; PID:AC25886.1; GSPDB:GN00023; CESP:F12F3.2
A/Experimental source: Strain Bristol N2; clone F12F3
C/Genetics:
A/Gene: CESP:F12F3.2
A/Map position: 5
A/Introns: 45/3; 90/3; 451/3; 509/1; 2313/3; 2341/3; 2378/3; 2414/2; 2453/3; 2474/2; 252

Query Match 5.7%; Score 133; DB 2; Length 2783;
Best Local Similarity 20.7%; Pred. No. 2.5;
Matches 94; Conservative 68; Mismatches 161; Indels 132; Gaps 22;

QY 28 VVLGKKGDTVELTCTAQSOKSIQFHWKNSNOIKILNGSFLTKGPSKLNDRADRSRLW 87
Db 1187 VVLTAETATATFTQSYANPAQVW-----LHNGKALQOTKSNYKTRLF 1231
QY 88 DQGNPFLIKLIKIEDSDTYICEVDQKEV-----QLLVFGLTANS----- 129
Db 1232 DQNTATLVIEVWDELGCTTAVANNOFGDVHTSAQLTISGEAKKTAASLPYIIEKLP 1291
QY 130 DTHLLQOSLTLTLESPPGSSP-----SVQCRSPR-----GKNIQSGKTLVSQ 173
Db 1292 KINVEG--ATLSIQADLNSGPPEVWVLKDNSELVSDRLOKCDGVNVQ---LVVRD 1345
QY 174 LELDQSGTWTCTVQLNQKVEFKIDIVLAFQKASSIVYKKEGEQVE-----S 222
Db 1346 VGLDEBGTYYITTAENEKIRQNTVEVS---TKSEVYEKKKKVEKQGGKKKPRFG 1402
QY 223 FPL---AFTVYKLT-----GSGELMMQAEARASSKSWITF-DLNKKEYSVK----- 264
Db 1403 LPRPSGASKTEQVTMAFDAPSEGPADSYEVRRCPPDQREBWSGCTSLSELEITGLPNT 1462
QY 265 ----RVTDPPQLQNGKULPLHLTLPQA---LPOYASGNLTALAEATGHLQEBVNLV 316
Db 1463 EYIFRVAQKKNQGGSEMSMTLKTASVQAPF-----TISPGK-----IIA 1507
QY 317 MRATQLOKULTCEVWGPSPKMLSLKLENKEAVYSREK-----PWVLA---PEA 365
Db 1508 NRDEFE--IAVEPSGPTP---SVKMYKENLQIVDEKIDVATTSSTSLNKSQEN 1561
QY 366 GMMQCLL-SDSGQVLLSNI-----KVLPTWSTPVH 395
Db 1562 GTFNCLIBNELGQASACQVITFNKPSLQSTPDH 1596

RESULT 23
T13669
neuromusculin - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 13-Aug-1999 #sequence revision 13-Aug-1999 #text_change 17-Nov-2000
C/Accession: T13669

R:Kantä, A.; Han, P. L.; Kim, Y. T.; Belten, H.
Neuron 11, 673-687, 1993
A:Title: Neuromusculin, a Drosophila gene expressed in peripheral neuronal precursors and in the CNS
A:Reference number: Z17697; MUID:94000831; PMID:8398154
A:Accession: T13669
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1011 <KAN>
A:Cross-references: EMBL:L23146; NID:G385073; PID:G385074; PIDN:AAA03750.1
C:Genetics:
A:Gene: nrm
A:Cross-References: FlyBase:FBgn0005629

Query Match	5.7%;	Score 132;	DB 2;	Length 1011;
Best Local Similarity	21.1%;	Pred. No. 0.75;		
Matches	83;	Conservative	68;	Mismatches 141;
				Indels 102;
				Gaps 20;

```

Qy      10 LLLVLQALALPAAT-OGNKKVVLGKKGTVELTCTASOKKSIDPH---WKNNOI--KIL 62
      ||| ||| : : : : : ||| : : : : : |||
Db      26 LVLVLCTALVDSSTAQVDTTISQGSQSVLPCFPVDAKCGKXLSLNFKGDRIAMILL 85

```

```

QY      63 GNQGSFLTKGSRKLNPRADSRSLWDCGNFPLIKLKIEDSTYICEVEDQKEVQLLV 122
      + : : | | : : : : | | | | | | : :
Db      86 GD-----SNVTSVKNKEFDERVTV-EQNPRYLVIKDLKIADIDYLCDT-----T 128

```

```

QY      123  FGLIANSPTHLQGGSLTTLTLESPPGSSPSVQCRRSPGCKNIQCGKTLTSLVQLLELDQSGTW 1822
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      129  FPIPEETCDN-FNGYRIELRVLP---TEVYILDAGKGRICKGVS--VGPMQERQSLKA 1822

```

```

Cy      163  |C|I|L|V|D|N|Q|K|A|Y|E|F|L|V|L|A|F|Q|M|S|S|I|V|K|R|E|G|O|V|E|F|S|P|L|A|T|Y|E|R|L|-T|S|S|G|E|L|W|Q| 241
          |||| : : : : : | : | : | : |
Db      183  TCTVANTRPPE-----VSMFGTKRLTTYSPTHDLVDGLYTSTLELDMT 227

```

228 LSRDLAQD---IECRVKSAAIQNVTVTKFSVDLQVRPTSIDINGVKHHTVQGSKVVLTC 284

285 DHHARPAVNUITWYNTTIISSGENEITEVRSKLSKSDGTFTQSEL-IFNATRFENDR 343

Db 344 VPRCEA-----ENIVLQIN-REKPI 362

RESULT 24
A88852
protein unc-22 [imported] - *Caenorhabditis elegans*
C. Sneideri; *Caenorhabditis elegans*

Science 282, 2012-2018, 1998
 R. anonymous, The C. elegans Sequencing Consortium.
 C1:Accession: A88852
 C1:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001

A₁Note: published errata appeared in *Science* 283, 35, 1999; *Science* 283, 2103, 1999.

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-6831 <STO>

C:Genetics:
A:Gene: unc-22
A:Map position: 4

```

Query Match          5.7%; Score 132; DB 2; Length 6831;
Best Local Similarity 21.8%; Pred. No. 9.8;
Matches 87; Conservativity 66; Mismatches 168; Total 255

```

QY 8 RHLVLQIALLPATQGNKVLGKGDVVELTCTASQKKSIFHWNKNSNQIKILGNQGS 67

Db	986	RHTEVV-----PMKSQ--KV--NESDLATLETVDNDKDAEYVMMHDKRIDIDGVK--	10322
Oy	68	FLTGKPSKLANDRAARSRLMDQGNFPLIIKLIKLETDSPTYICEVEODKEEVOULTFVg--	124
Db	1033	-----FKVESNRKR-----LIIINGARIEDGEVKECTTKDKDRTHAQLIVDAKAK	1077
Oy	125	-LTRANSDTHLQOGSLTFLTESPPGSSPSVOCSPRGKNIQ--GgK-----TlS	170
Db	1078	FIVALKDTEVIEKODVILMCOTKDTKTPGAIWFRN--GKQISMPGKGFEIROSNGRHTLTK	1135
Oy	171	VSQLELDSGWTCTVUNQOKKAEFKIDIVLAFQKASSIYVK-----KEGEVSEESFP	224
Db	1136	IGKIEMBADYEL-----DQAGRGSGCNVTVLEAEKRPILNMWKEPKIEAAGBECVAVKP	1191
Oy	225	LAFVETKLTSGGELIMQOABERASSSKSWITPDLKNAEYSVKRYODPRLQMGKPLPHLTL	284
Db	1192	FQI---KSTRRGD-----PKQIILKNGKRPIDEBEKK--VEVILIKDVAELVFK-----	1235
Oy	285	QALPQVAGSGNLTALAEAKTGKJHDEVNLVVM-----BATOLQKNUJCE---VMGPT	334
Db	1236	---NPOGLADYGKMALEIGNSAGTALAPFELFVVDKPKRPGRPLETKNVTAEGLDVLWGTP	1292
Oy	335	SPKMLSLK--LENKE-----AKVSK-----REKPVVNLNPEAGMMOQLSD	374
Db	1293	DPDEGAPVAKVYIIEMQGRSGNMAKVGETKGTDFKVDLKEHGHEKFRVAK--INECGLSD	1351
Oy	375	--SGOVILLESIKVILPTWSTPVIHR	397
Db	1352	PLTESVLAKN---PYGVPGKK	1371

RESULT 25
S57242

N-Alternate names: myosin-regulating protein
N-Contains: protein kinase (EC 2.7.1.-)
CISpecies: *Caenorhabditis elegans*

R. Benjan, G.M.; L'Hernault, S.W.; Morris, M.E.
submitted to the EMBL Data Library, February 1993
a. <http://www.ebi.ac.uk/EMBL/seq/submit/>

A: Reference number: S57242
A: Accession: S57242
A: Molecule type: DNA
A: Residues: 16830 <BBN1>

A/cross-references: EMBL:U031
 A/Experimental source: var. Bristol
 R.Bentley, G.
 submitted to the EMBL Data library, November 1989

A:Accession: S07571
A:Molecule type: DNA
A:Residues: 792-6839 <BEN2>

A: Experimental source: var. Bristol
R. Benian, G.M.; Kiff, J.E.; Neckelmann, N.; Moerman, D.G.; Waterston, R.H.
Nature 342, 45-50, 1989

A1:Reference number: S06797, MUID:90044042, PMID:2812002
A2:Accession: S06797
A3:Status: nucleic acid sequence not shown

A: Experimental source: Var. Bristol
P: Reptan G M . l'Hernault S W . Morris M F
A: Residues: 806-1175,1178-1998, 'Y', 2000-3040, 'I', 3042-3335, 'I', 3337-35693; 5696-6
A: Cross-references: EMBL: X15423

Genetics 134, 1057-1104, 1993
 A1Title: Additional sequence complexity in the muscle gene, unc-22, and its encod
 A1Reference number: S57218; MUID:93387664; PMID:8397135
 A1Accession: S57218

A;Molecule type: DNA
A;Accession: 55/218

A:Residues: 2-99;108-194,'Q',196-206;374-468;658-753 <BBN4>
A:Experimental source: var. Bristol
R:White, S.
Submitted to the EMBL Data Library, May 1996
A:Reference number: Z20442
A:Accession: T27934
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 'MGIPGKKCKQ',19-6839 <W12>
A:Cross-references: EMBL:Z73897; PIDN:CAA98064.1; GSPDB:GN00022; CESP:ZK617.1a
A:Experimental source: clone ZK617
R:Harris, B.
Submitted to the EMBL Data Library, May 1996
A:Reference number: Z20458
A:Accession: T28030
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 'MGIPGKKCKQ',19-6839 <W12>
A:Cross-references: EMBL:Z73899; PIDN:CAA98081.1; GSPDB:GN00022; CESP:ZK617.1a
A:Experimental source: clone ZK829
C:Comment: Lack of unc-22 leads to a constant twitching of the body muscles.
C:Genetics:
A:Gene: unc-22; CESP:ZK617.1a
A:Map position: 4
A:introns: 18/3; 69/3; 143/2; 176/3; 264/2; 387/3; 413/2; 471/1; 516/3; 550/3; 582/3; 60152/3; 6691/3; 6776/1; 6808/3
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;
C:Keywords: ATP, autophosphorylation, duplication, muscle, phosphotransferase, serine/th
F:806-898,899-990,991-1083,1084-1175,1178-1273,1474-1567,1770-1864,2066-2158,2358-2450,2
96-5750,6263-6356,6386-6478,6541-6635,6649-6742,6745-6838/Region: motif 2
F:1274-1372,1373-1473,1568-1670,1671-1769,1865-1964,1965-2065,2159-2258,2259-2357,2451-2
23,4215-4313,4314-4415,4416-4516,4612-4710,4711-4811,4908-5009,5010-5109,5110-5210,5399-
F:5940-6197/Domain: protein kinase homology <KIN>
F:5944-5956/Region: protein kinase ATP-binding motif
F:5971/Active site: Lys Abacatus predicted

Query Match 5.7%; Score 132; DB 2; Length 6839;
Best Local Similarity 21.8%; Pred. No. 9.8;
Matches 97; Conservative 66; Mismatches 168; Indels 114; Gaps 23;

QY 8 RHLLVQLALLPATQGNKVVLGKKGDTVELTCTASQKSIQFHWKNSNOIKILGNQGS 67
DB 994 RHTFVV-----PMKSC--KV---NESDLATLETVDNDKDAEVMWMDGKRIDIDGVK-- 1040
QY 68 FLTKGPKSLNDRADSRSLMDQGNPFLIKLIKIEDSTYICVEDQKEVQLVFG--- 124
DB 1041 -----FKVSSSNRRKR-----LINGARIEDHGEYCTTDDRTMAQLIVDAK 1085
QY 125 -LTRANSDTHLQGSILTLTLESPGSSPSVOCSPRGKNIQ---GSK-----TIS 170
DB 1086 FIVALKDTEVIEKDVTLMCQTKDTGTFGRN--GKQISSMPGSKFETQSRNGHTLTK 1143
QY 171 VSQLELDQSGTWTCTVVLQNKVEFKIDIVLAFQKASSIVYK-----KEGEVFSFSP 224
DB 1144 IKIEMENADYVEI---DQAGLRGSCNVTLEAEKRPILNMKPKKIEAKGEPGVKVP 1199
QY 225 LAFVTEKLTGSGELMWQAEARASSKSWITPDLKXKVSVKVNTQDPKQIMGKKLPLHLTL 284
DB 1200 FQI---KGTTRGD-----PKAQLIKNGKPIDEEMKCL-VEVITIKDDVAEIVFK----- 1243
QY 285 PQALPQVAGSNLTLALEAKTGKTHQEVNLVVM-----RATOLQKNLTCE---VMGPT 334
DB 1244 ---NPQLADTGKMALELGNAGTALAPFLFKDKRPKPKPLETNTVTAEGLDLVWGTP 1300
QY 335 SPKMLSLK---LENKE-----AKYSK-----REKPVWVNLPEAGMOCCLSD 374
DB 1301 DPDEGAPVKAAYIIEHQEGRSGNWAKVGETKTDPKVKDLKEHGEYKFRVKA-LNECGSLD 1359
QY 375 ---SGQVLLSNTKVLPTWSTPVHPR 397
DB 1360 PLTGESVLAKN-----PYGVGPKPK 1379

RESULT 26
T27935
hypoetical protein ZK617.1b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
C:Accession: T27935; T28031
R:White, S.
Submitted to the EMBL Data Library, May 1996
A:Reference number: Z20442
A:Accession: T27935
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7160 <W12>
A:Cross-references: EMBL:Z73897; PIDN:CAA98065.1; GSPDB:GN00022; CESP:ZK617.1b
A:Experimental source: clone ZK617
R:Harris, B.
Submitted to the EMBL Data Library, May 1996
A:Reference number: Z20458
A:Accession: T28031
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7160 <W12>
A:Cross-references: EMBL:Z73899; PIDN:CAA98082.1; GSPDB:GN00022; CESP:ZK617.1b
A:Experimental source: clone ZK829
C:Genetics:
A:Gene: CESP:ZK617.1b
A:Map position: 4
A:introns: 10/3; 61/3; 135/2; 168/3; 256/2; 379/3; 405/2; 463/1; 508/3; 542/3; 574/3; 593067/1; 5141/3; 3469/1; 6473/3; 7012/3; 7097/1; 7128/3
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

Query Match 5.7%; Score 132; DB 2; Length 7160;
Best Local Similarity 21.8%; Pred. No. 10;
Matches 97; Conservative 66; Mismatches 168; Indels 114; Gaps 23;

QY 8 RHLLVQLALLPATQGNKVVLGKKGDTVELTCTASQKSIQFHWKNSNOIKILGNQGS 67
DB 1315 RHTFVV-----PMKSC--KV---NESDLATLETVDNDKDAEVMWMDGKRIDIDGVK-- 1361
QY 68 FLTKGPKSLNDRADSRSLMDQGNPFLIKLIKIEDSTYICVEDQKEVQLVFG--- 124
DB 1362 -----FKVSSSNRRKR-----LINGARIEDHGEYCTTDDRTMAQLIVDAK 1406
QY 125 -LTRANSDTHLQGSILTLTLESPGSSPSVOCSPRGKNIQ---GSK-----TIS 170
DB 1407 FIVALKDTEVIEKDVTLMCQTKDTGTFGRN--GKQISSMPGSKFETQSRNGHTLTK 1464
QY 171 VSQLELDQSGTWTCTVVLQNKVEFKIDIVLAFQKASSIVYK-----KEGEVFSFSP 224
DB 1465 IKIEMENADYVEI---DQAGLRGSCNVTLEAEKRPILNMKPKKIEAKGEPGVKVP 1520
QY 225 LAFVTEKLTGSGELMWQAEARASSKSWITPDLKXKVSVKVNTQDPKQIMGKKLPLHLTL 284
DB 1521 FQI---KGTTRGD-----PKAQLIKNGKPIDEEMKCL-VEVITIKDDVAEIVFK----- 1564
QY 285 PQALPQVAGSNLTLALEAKTGKTHQEVNLVVM-----RATOLQKNLTCE---VMGPT 334
DB 1565 ---NPQLADTGKMALELGNAGTALAPFLFKDKRPKPKPLETNTVTAEGLDLVWGTP 1621
QY 335 SPKMLSLK---LENKE-----AKYSK-----REKPVWVNLPEAGMOCCLSD 374
DB 1622 DPDEGAPVKAAYIIEHQEGRSGNWAKVGETKTDPKVKDLKEHGEYKFRVKA-LNECGSLD 1680
QY 375 ---SGQVLLSNTKVLPTWSTPVHPR 397
DB 1681 PLTGESVLAKN-----PYGVGPKPK 1700

RESULT 27
A41060
neural cell adhesion molecule LI precursor - human
N:Alternate names: LiCAM
C:Species: Homo sapiens (man)

Matches 85; Conservative 68; Mismatches 195; Indels 88; Gaps 18;

QY 33 KGDVETLTCTASQKSIQFHWKNSNQIKILNGQSFLLTGKPSKLNDRADRSRLMDQGNF 92
 Db 3495 ENDITMDCGVTSRPLPSISM-----FRGDKPYLYLD---RYSISPDGSH 3536

QY 93 PLIINKLKIEDSDTYICEVEDQ--KEEVLVFGLT-ANSDTHLLQGQSL-----TLTLE 144
 Db 3537 -ITINKAKLSDGKXICRASNEAGTSDIDLILKILVPRKIDKSNIIIGNPLAVARTIYLE 3595

QY 145 SPFGSSP-----SVQCRSPRGKNIQGGKTLVSQLELDQSGTCTVLOQNKV 193
 Db 3596 CPISGIPQPDVITWTKMGMDIMTDSRVLLAQNNEFFGLENVQVTDQGRYTCATNRGGA 3655

QY 194 EFKIDIVLAFQK---ASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMWQAERASSK 249
 Db 3656 SHDFSLDVLSPEFDHGTQPTIKREGDTITLTCPKLAEDIADQVMDVSWTKDSRA--- 3712

QY 250 SWITFDL-KNKEVSVKRVTDPRKIQMGKPLHLTLPQALPQYAGSGNLTALAEKTKL 308
 Db 3713 --LDGDLTDNDVDS-----DDGRK---LTISQASLENNAGLYTCIALNRAGEASL 3756

QY 309 HOEVLVWMRATQLOKN-----LTCEWGGTSPKLMSLK-----LENKEAK 350
 Db 3757 EFKIEILSPVIDISRNDVQPVAVNQPTIMRCVATGHFPFSIKM-LKNGKEVTDENIR 3815

QY 351 VSKREKPVWVNLNPE---AGMOCCLL-SDSGVLLLESNIKVLPTWSTPVPRASALPAPT 406
 Db 3816 IVEGQGVQLRLTSDSHAGKSCVAENDAGVKELE---MVLDFTPPVVSVKSDNPICAL 3872

QY 407 GSALPDPTASALPDP 422
 Db 3873 GETITLFCNASGNPYP 3888

RESULT 29

T43290
 hemiscenlin precursor - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #text_change 18-Feb-2000
 C:Accession: T43290; T20993; T24734
 R:Vogel, B.E.; Hedgecock, E.W.
 submitted to the EMBL Data Library, June 1998
 A:Description: Hemiscenlin is required for hemidesmosome mediated cell adhesion and germ-
 A:Accession: T43290
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-5198 <VOG>
 A:Cross-references: EMBL:AF074901; PIDN:AA26792.1
 R:Sulston, J.
 submitted to the EMBL Data Library, December 1994
 A:Accession: T20993
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-5198 <WID>
 A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:FL5G9.4b
 A:Experimental source: clone F15G9
 R:Kerhan, J.
 submitted to the EMBL Data Library, December 1994
 A:Reference number: Z19355
 A:Accession: T24734
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-5198 <W12>
 A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:FL5G9.4b
 A:Experimental source: clone T09B9
 C:Genetics:
 A:Gene: him-4; F15G9.4b
 A:Map position: X
 A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
 ; 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1

1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 5100/

Query Match 5.6%; Score 131; DB 2; Length 5198;
 Best Local Similarity 19.5%; Pred. No. 7.9;
 Matches 85; Conservative 68; Mismatches 195; Indels 88; Gaps 18;

QY 33 KGDVETLTCTASQKSIQFHWKNSNQIKILNGQSFLLTGKPSKLNDRADRSRLMDQGNF 92
 Db 3495 ENDITMDCGVTSRPLPSISM-----FRGDKPYLYLD---RYSISPDGSH 3536

QY 93 PLIINKLKIEDSDTYICEVEDQ--KEEVLVFGLT-ANSDTHLLQGQSL-----TLTLE 144
 Db 3537 -ITINKAKLSDGKXICRASNEAGTSDIDLILKILVPRKIDKSNIIIGNPLAVARTIYLE 3595

QY 145 SPFGSSP-----SVQCRSPRGKNIQGGKTLVSQLELDQSGTCTVLOQNKV 193
 Db 3596 CPISGIPQPDVITWTKMGMDIMTDSRVLLAQNNEFFGLENVQVTDQGRYTCATNRGGA 3655

QY 194 EFKIDIVLAFQK---ASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMWQAERASSK 249
 Db 3656 SHDFSLDVLSPEFDHGTQPTIKREGDTITLTCPKLAEDIADQVMDVSWTKDSRA--- 3712

QY 250 SWITFDL-KNKEVSVKRVTDPRKIQMGKPLHLTLPQALPQYAGSGNLTALAEKTKL 308
 Db 3713 --LDGDLTDNDVDS-----DDGRK---LTISQASLENNAGLYTCIALNRAGEASL 3756

QY 309 HOEVLVWMRATQLOKN-----LTCEWGGTSPKLMSLK-----LENKEAK 350
 Db 3757 EFKIEILSPVIDISRNDVQPVAVNQPTIMRCVATGHFPFSIKM-LKNGKEVTDENIR 3815

QY 351 VSKREKPVWVNLNPE---AGMOCCLL-SDSGVLLLESNIKVLPTWSTPVPRASALPAPT 406
 Db 3816 IVEGQGVQLRLTSDSHAGKSCVAENDAGVKELE---MVLDFTPPVVSVKSDNPICAL 3872

QY 407 GSALPDPTASALPDP 422
 Db 3873 GETITLFCNASGNPYP 3888

RESULT 30

S04663
 T-cell receptor gamma chain - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 28-Feb-1990 #sequence revision 28-Feb-1990 #text_change 21-Jan-2000
 C:Accession: S04663
 R:Okada, A.; Bank, I.; Rogozinski, L.; Takihiro, Y.; Mak, T.W.; Chess, L.; Alt, F.W.
 J. Exp. Med. 168, 1481-1486, 1988
 A:Title: Structure of the gamma/delta T cell receptor of a human thymocyte clone.
 A:Reference number: S04663; MUID:89010543; PMID:2844954
 A:Accession: S04663
 A:Molecule type: DNA
 A:Residues: 1-304 <OKA>
 A:Cross-references: EMBL:X15018
 A:Note: The authors translated the codon AAC for residue 72 and AAT for residues 102 and
 C:Genetics:
 A:Introns: 15/1; 133/3
 C:Superfamily: immunoglobulin V region; immunoglobulin homology
 C:Keywords: T-cell receptor
 F:158-223/Domain: immunoglobulin homology <IMM>

Query Match 5.6%; Score 130.5; DB 2; Length 304;
 Best Local Similarity 21.3%; Pred. No. 0.19;
 Matches 74; Conservative 59; Mismatches 128; Indels 87; Gaps 15;

QY 8 RHLLVQLALPAAQGNKVVLLGKSDT-----VELTCTASQKSIQFHW----- 53
 Db 2 RWAIVLVLAFLSPASQSSNLEGGTKSVTRPRSSAEITCDLYINAFYIMYLAHQEKA 61

QY 54 -----KNSNQIKILG-NQGSFLTGKPSKLNDRADRSRLMDQGNFPLIKLKIED 103
 Db 62 PQRLLIYDVNSKVDLESGLSPKYYHTP-----KRWSN-----ILLRLIEND 107

QY 104 SDTYICEVEDQKEEVQLVFLGLTANSDTHLLQGQSLTLT--LSPFGSSPSVQCRSPRG 160

```

Db      108 SGVYVCATMDRQK---KLFG-----SGFTLVVTDKQDADAVSPKFTFLP----- 150
QY      161 KNIGQKTLVSQLELDSDSTWCTYLQNKQKVEFKIDIVLAFQ--KASSIYKKEGEG 218
Db      151 -----SIATKTKQKATGYLCL-----SKFPFDIKIKIMQEKSKNTIIGSEGN 195
QY      219 -----VEFSPLAFVETLTSQGLMWAERASSKSWITFDLKNKEVSRYKVTQDPK 271
Db      196 MKTNDTYMKFSW-LTPPEESLDKSHRC---IVHNNKNGIDDEIIFPPIKTDVTVDPK 251
QY      272 LQWKKLPLTLTPQALPOYAGSGNLTALAEATKGLHDEVNLVWRA 319
Db      252 YNYSKQANDVITM-DPRDMSKQANDTLLQLNTSAVYVYVLLLLKS 298

RESULT 31
myosin-light-chain kinase (EC 2.7.1.117), 210K, nonmuscle - chicken
N:Contains: myosin-light-chain kinase, 108K, smooch muscle; telokin
C:Species: Gallus gallus (chicken)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C/Accession: S68235; A37099; B44389; S28227; S78216; A35093; A25810; S11652
R:Walterson, D.M.; Collinge, M.; Lukas, T.J.; Van Eldik, L.J.; Bitukov, K.G.; Stepanova,
FEBS Lett. 373, 217-220, 1995
A>Title: Multiple gene products are produced from a novel protein kinase transcription
A/Reference number: S68235; MUID:96033976; PMID:7589469
A/Accession: S68235
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-1906 <EMBL>
A/Cross-references: EMBL:X52876; NID:9992992; PIDN:CAA37056.1; PID:9992993
R:Shoenaker, M.O.; Lau, W.; Shattuck, R.L.; Kwiatkowski, A.P.; Matrisian, P.E.; Guerra-S
J. Cell Biol. 111, 1107-1125, 1990
A>Title: Use of DNA sequence and mutant analyses and antisense oligodeoxynucleotides to
activity.
A/Reference number: A37099; MUID:90361738; PMID:2202734
A/Accession: A37099
A/Molecule type: mRNA
A/Residues: 649-1906 <SHO>
A/Cross-references: EMBL:X52876
R:Collinge, M.; Matrisian, P.E.; Zimmer, W.E.; Shattuck, R.L.; Lukas, T.J.; Van Eldik, I
Mol. Cell. Biol. 12, 2359-2371, 1992
A>Title: Structure and expression of a calcium-binding protein gene contained within a c
A/Reference number: A44389; MUID:92236611; PMID:1373815
A/Accession: B44389
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1695-1906 <COL>
A/Cross-references: GB:M88284; NID:g212237; PIDN:AAB5767.1; PID:g212238
A/Accession: A44389
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1750-1906 <CO2>
A/Cross-references: GB:M88283; NID:g211371; PIDN:AAA48647.1; PID:g211372
R:Yoshikai, S.I.; Ikebe, M
Arch. Biochem. Biophys. 299, 242-247, 1992
A>Title: Molecular cloning of the chicken gizzard telokin gene and cDNA.
A/Reference number: S28227; MUID:93073972; PMID:1444462
A/Accession: S28227
A/Molecule type: mRNA
A/Residues: 1750-1906 <YOS>
A/Cross-references: EMBL:M6655; NID:g212744; PIDN:AAA49083.1; PID:g212745
A/Accession: S78216
A/Molecule type: DNA
A/Residues: 1750-1906 <YOM>
A/Cross-references: EMBL:M69687
R:Olson, N.J.; Pearson, R.B.; Needleman, D.S.; Hurwitz, M.Y.; Kemp, B.E.; Means, A.R.
Proc. Natl. Acad. Sci. U.S.A. 87, 2284-2288, 1990
A>Title: Regulatory and structural motifs of chicken gizzard myosin light chain kinase.
A/Reference number: A35093; MUID:90192792; PMID:2315320
A/Accession: A35093
A/Status: preliminary

```

```

A/Molecule type: mRNA
A/Residues: 935-1438, 'Q', 1440-1906 <OLS>
A/Cross-references: GB:M31048; NID:g212660; PIDN:AAA9069.1; PID:g212661
R:Guerrero Jr., V.; Russo, M.A.; Olson, N.J.; Putkey, J.A.; Means, A.R.
Biochemistry 25, 8372-8381, 1986
A>Title: Domain organization of chicken gizzard myosin light chain kinase deduced from
A/Reference number: A25810; MUID:87157587; PMID:3030394
A/Accession: A25810
A/Molecule type: mRNA
A/Residues: 1258-1438, 'Q', 1440-1906 <GUE>
C/Genetics:
A/Intons: 1735/3; 1779/1; 1819/1
C/Superfamily: myosin-light-chain kinase, nonmuscle; fibronectin type III repeat homolo
C/Keywords: alternative initiators; ATP; calmodulin binding; phosphoprotein; phosphotran
F:542-599/Domain: immunoglobulin homology <IMM1>
F:935-1906/Domain: myosin-light-chain kinase, 108K, smooch muscle (from 5.5kb transcript
F:1098-1158/Domain: immunoglobulin homology <IMM2>
F:1451-1708/Domain: protein kinase homology <KIN>
F:1459-1467/Region: protein kinase ATP-binding motif
F:1750-1906/Product: telokin (kinase-related protein KRP) (from 2.7 kb transcrip
F:1808-1869/Domain: immunoglobulin homology <IMM3>

Query Match      5.6%; Score 130; DB 1; Length 1906;
Best Local Similarity 20.8%; Pred. No. 2.4;
Matches 82; Conservative 58; Mismatches 142; Indels 112; Gaps 15;

QY      89 QGNFPIITNKLIEDSDTYICEVD---QKEVQLVFGLTANSPTHLQGSITLTLE 144
Db      81 RGFISLVIKGVQEGSGKTYCEANDGVQVTELVTEG---NS---LKKYSIPSSAK 133

QY      145 SPRG---SSPSVOCR-----SPRNGIGGKT----- 168
Db      134 TREGRLSVPVEHRDSIMGESPPKATKRNRYVVBGGTGRFSKITGRRPQVWTAKD 193

QY      169 -----LSVSOLELDSDSTWCTYLQNKQKVEFKIDIV----- 201
Db      194 IHLQNERFMFEKTIQVLEIQNVGLADAGITCTVNVASAKASVSELVTOGPDXTD 253

QY      202 -----LA-----FOKASSIVYKKEGVEFSPLAFYVEKLTSGSELW 240
Db      254 HAOPLCMPKPTTLTKALINSDFKQATNSNGIAKEIKSTSEL-VWETDRLSAKKEFY 312

QY      241 QABRASSKSWITFDLKNKEVSVKRTOOPKQLQWKKLPLHLTLPOALPOYAGSGNLTALA 300
Db      313 TSEBAKQKQGNQANNAVPLQESRGTGKPPVLQKTSSTITLQAVVAQDEPPAEPQTFI 372

QY      301 LEAKTGKLTQEVNLVWRATQLQKNLTCEVWGPTSPKMLSLKEN-----KEAKYSKE 355
Db      373 RQAEHRK--RTVQPLMTTITTOENPSLT-----GVSPR---SRETNRAGVRSVBEKKE 423

QY      356 KPVWNLNPEAGMWQCLLSDSQ-VILLESNIKVL 388
Db      424 -PLGIPPGFESRPGQLEASGEIRKFSKVSQGP 456

RESULT 32
T42633
connectin/titin - chicken (fragment)
C/Species: Gallus gallus (chicken)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C/Accession: T42633
R:Ujima, H.; Onitsuka, H.; Kawamura, Y.; Kume, H.; Murayama, T.; Abe, H.; Kimura, S.; Ma
Biochem. Biophys. Res. Commun. 223, 160-164, 1996
A>Title: A 11.5-kb 5'-terminal cDNA sequence of chicken breast muscle connectin/titin re
A/Reference number: Z22221; MUID:96254045; PMID:8660363
A/Accession: T42633
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-4162 <YAJ>
A/Cross-references: EMBL:D83390; NID:g1513029; PIDN:BAAL1908.1; PID:g1513030
A/Experimental source: breast muscle
C/Keywords: skeletal muscle

```

```

Query Match          5.6%; Score 130; DB 2; Length 4162;
Best Local Similarity 22.4%; Pred. No. 6.8;
Matches 103; Conservative 55; Mismatches 159; Indels 142; Gaps 24;

QY 11 LVLQLALLPAATQGNKVVLGKKGVLTCTASQKKSIOFW-KNSNOI-----KILG 63
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 3551 LTVLEPAVAVEKPGPVKVT---AGDSCTLECTVDGTPELTARWFGDGLSTDHAKYKI-- 3605
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 64 NGGSFLLTK--GPSKLNDRADSRSLMDQGNFPLITKIEDSDPYICVEVDQKEVQLL 121
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 3606 ---SFPNKSGLKILN-----AGLEDSEGYTFPEVKNSGKSSCTASLQVSDR----- 3649
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 122 VEGLTANSDTHLQGSQSLTLDE---SPGSSPVQCR--SP-----RGKNIQGGK 167
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 3650 -----IMPPSFRKLKETGYGLGSSAVLECKVYGSPPILVSWFHDGGEITSD 3697
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 168 -----TLSSQLELDSDSGTWTCTVLOKKNVEKIDIVLAPKASIIYVKKK- 215
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 3698 KYQATLTDTNCTGLKNGQLQESDMGTSTATVAVSGDCS--AFLSVAREPPSPVKKEP 3754
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 216 -----GEOVEFSFPLAFTVEKLTGSGELW---QAERASSKSWITPDLKKEVSVKRV 267
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 3755 FNVLSGENTFTT-----SLVKGSPLELVNMFPGSIEFLAPGHKCNITL----- 3796
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 268 QD--PKLQNGKKLPLHLTLPOALPOYAGSGNLTALAEAKTGKLHGEVNL-----VYWR 318
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 3797 QDSVALELFDVQPLQ-----SGDYTCQVSNBAGKISCTTHLFLVKEPAKFWK 3844
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 319 ANQQLQ---KNLTCEWGTSPKLMSTLKEKKEAKVSKREKPVWLVNEA----- 365
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 3845 VNDLSVEKKNLILECTYTGTPISVATWK--KNGVILHSGKCSITTTETSAIIIEPNS 3901
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 366 -----GMQOC-LISDSQVLTLESNIKVL--PTWSTFVHP 396
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 3902 KLEDGQVYCHLENDSGQDNCGALTILEPFFVTPLEP 3940
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 33
T32735
telomerase-associated protein component 1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T32735
R:Nakayama, J.; Saito, M.; Nakamura, H.; Matsura, A.; Ishikawa, F.
Cell 88, 875-884, 1997
A:Title: TLPI: a gene encoding a protein component of mammalian telomerase is a novel me
A:Reference number: Z21217; MUID:97236507; PMID:9118230
A:Accession: T32735
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2629 <NAK>
A:Cross-references: EMBL:U89282; NID:91932816; PID:91932817; PIDN:ABS1690.1
A:Experimental source: strain Fischer 344
A:Genetics:
A:Gene: TLPI
C:Function:
A:Description: modification of telomerase-associated protein component 1 may regulate te

```

```

QY 139 LTL-----TLESPGSSPVQCRSPRGK-----IQGKTLVSOLELQDS 179
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 2395 LIIMKEDEVELQMPRGSTPSSICRVAHSSILCTSKDYGLFYIQGNGSSLSILEGES 2454
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 180 GTWCTCTVLOKQKVEFK-----IDIVLAPQKASIIYVKKKEQVEFSFPLATV 229
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 2455 GKF-----EKTIDFNLNLNPNGPSVSLTQAEPEGSSSL-----CATSGMLMNL 2500
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 220 EKLGSGB-----LWQAERASSK-----SWITFDLKNKEVSVKRVQDP 270
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 2501 SECTPEGEWVDVNTWQKSRNPKSRTTPEQDSSPGLFCMDSVW-----BPTHILAROCK 2553
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 271 KLOMGKKLPLHLTLPOALPOYAGSGNLTALAEAKTGKLHGEVNLVWPARATQLOKNTLC-E 329
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 2554 KIHGSVTALHV-----LP-----GLVTASEDRDVKLMERSMQLGLFRCCEPVSCLE 2603
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 330 VWGFTSPKLMSTL 342
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 2604 PMWEPSPQLQAV 2616
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 34
S18252
heparan sulfate proteoglycan - mouse
N:Alternate names: perlecan
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
A:Accession: S18252; A31917; E31917; S66460
R:Noonan, D.M.; Fuller, A.; Valente, P.; Cai, S.; Horigan, E.; Sasaki, M.; Yamada, Y.; Ha
J. Biol. Chem. 266, 22939-22947, 1991
A:Title: The complete sequence of perlecan, a basement membrane heparan sulfate proteogly
adhesion molecule.
A:Reference number: S18252; MUID:92078153; PMID:1744087
A:Accession: S18252
A:Molecule type: mRNA
A:Residues: 1-3707 <NOO>
A:Cross-references: EMBL:M77174; NID:9200295; PIDN:AAA39912.1; PID:9200296
R:Noonan, D.M.; Horigan, E.A.; Ledbetter, S.R.; Vogeli, G.; Sasaki, M.; Yamada, Y.; Hase
J. Biol. Chem. 263, 16379-16387, 1988
A:Title: Identification of cDNA clones encoding different domains of the basement membra
A:Reference number: A92680; MUID:85034110; PMID:2972708
A:Accession: A31917
A:Molecule type: mRNA
A:Residues: 940-1601 <NO2>
A:Cross-references: GB:J04054; NID:9200252; PIDN:AAA39899.1; PID:9200253
A:Accession: B31917
A:Molecule type: mRNA
A:Residues: 1870-2600 <NO3>
A:Cross-references: GB:J04055; NID:9200300; PIDN:AAA39912.1; PID:9200301
R:Schulze, B.; Mann, K.; Battierutta, R.; Wiedemann, H.; Timpl, R.
Eur. J. Biochem. 231, 551-556, 1995
A:Title: Structural properties of recombinant domain III-3 of perlecan containing a glob
A:Reference number: S66460; MUID:95377282; PMID:7649134
A:Accession: S66460
A:Molecule type: Protein
A:Residues: 1272-1274, 'X', 1276, 'X', 1278-1279 <SCH>
A:Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repe
C:Keywords: glycoprotein
F:1199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:1285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:764-811/Domain: laminin-type EGF-like homology <LEG>
F:1159-1206/Domain: laminin-type EGF-like homology <LEG7>
F:1153-1610/Domain: laminin-type EGF-like homology <LEG8>
F:1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F:3163-3198/Domain: EGF homology <EGF>
F:3270-3423/Domain: laminin G repeat homology <LG2>
F:3464-3492/Domain: EGF homology <EGF7>
F:1256,1891,2336,2394,2427/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

Query Match 5.5%; Score 127; DB 2; Length 3707;

Best Local Similarity 19.3%; Pred. No. 9.2;
Matches 100; Conservative 66; Mismatches 185; Indels 168; Gaps 22;

QY	32	KGGTVELTCA	SSKKSIQFHWKNSNQIKILGNQGSFLTGSPSLYNDRADSRRLMDQGN	91
Db	2447	KMGDITL	ECTSSSEPRSSPRMTR-----LGT-----PVKLEPRMGLKMN-----SH	2488
QY	92	FPLLIKNL	KIIBDSPTYICEVED---OKEEVOLLFGLTANSDHTLLQGSLLTLLTSPR	147
Db	2489	AMLKIAIVK	PSDADATGVYCSQAQNALGTAKQVELIVDTGTAVPTQYUVESELTLLA--	2546
QY	148	GSSPSVQCR	-----PRGNIGQGTLSVSOLEDDSGTWTCTVVLQ	190
Db	2547	GHTVTLT	HCASAGNRPRTTHMSKLRAPRYWQHRIE-GNTLVIPVAAQDSGYICNATNSA	2605
QY	191	KKVEFKID	IVVLAFOKASSI---VYKKEGEQVEFSF-----PLAF-----	227
Db	2606	GHTBATV	LVNHESPRTATIPENTSAQGNLVQGLSLAHGTRPFLTYQMSLVGVLPBKAV	2665
QY	228	-----	-----TYEKLTSGGSLMWOABASSKSKYITFDLKNKSVSKRY	266
Db	2666	VRNQLRL	LEFTVPBDSGRYRCQVNSNRVGSAAFAQULVWQSSSNLPTDTSIPGSGTPTVQV	2725
QY	267	TODEPLQ	---MGKELPLHLLTLPQA---LPOYASGNLTLLAEKGTCLHQEVLVYMRA	319
Db	2726	T--PQLER	NRIGASVETFCAPVNRNERTGHLRYLKHGGQLPRHNSVQDG-----VLR	2774
QY	320	TOLOK	---LTCSEVBR-----TSPKMLSLKLENKEAVU-----	351
Db	2775	QNLRON	CSGTVYCSAHGFWGQAOATLQILVALPSVULNVTSHSVUUGHSVEFECLAL	2834
QY	352	S-SKEKPVV	-----LND-----EAGMOCCLSDS-----GGVLESN	383
Db	2835	GDRPQV	TKSVKGVGHLRPGIVQSGTIIIRIAVELADKQYUHCATNNAAGTTQSHVLL--L	2892
QY	384	IKVLPTW	STVPHRPSALRAPRTGSALPDPQTASALRDP	422
Db	2893	VOALR	POISTREIRV-----PAGSAAVFPCMAAGUPTR	2925

RESULT 35
 NUSNG
 neural cell adhesion molecule 1 precursor, GPI-anchored splice form - mouse
 N/Alternate names: NCAM-120
 C/Species: Mus musculus (house mouse)
 C/Date: 31-Mar-1993 #sequence revision 31-Mar-1993 #text_change 31-Dec-2000
 C/Accession: A29673; S00382; A44290
 R/Bartshals, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaix, J.C.; Hirsch, M.R.; Fontec
 EMBO J. 6, 907-914, 1987
 A>Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,000
 A/Reference number: A29673; MUID:87246524; PMID:3595563
 A/Accession: A29673
 A/Molecule type: mRNA
 A/Residues: 1-725 <BAR>
 A/Cross-references: EMBL:Y00051; NID:G53342; PIN:CA68263.1; PID:G53343
 R/Bartshals, J.A.; Chaix, J.C.; Steinetz, M.; Goriadis, C.
 EMBO J. 7, 625-632, 1988
 A>Title: Differential splicing and alternative polyadenylation generates distinct NCAM t
 A/Reference number: S00382; MUID:88283628; PMID:3396534
 A/Accession: S00382
 A/Molecule type: DNA
 A/Residues: 642-656, 'D', 658-725 <BA2>
 A/Cross-references: EMBL:X07195
 R/Rougon, G.; Marshak, D.R.
 J. Biol. Chem. 261, 3396-3401, 1986
 A>Title: Structural and immunological characterization of the amino-terminal domain of π
 A/Reference number: A44290; MUID:86140120; PMID:3512556
 A/Accession: A44290
 A/Molecule type: protein
 A/Residues: 20-36 <ROU>
 A/Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
 A/Genetics: Several forms of NCAM are produced by alternative splicing. See also PIR:JUMS

A:Gene: NCAM
A:Map position: 9
A:introns: 701/1
C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immunoglobulin domain
C:Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membrane protein
F:1-19/Domain: signal sequence #status predicted <Sig>
F:34-98/Domain: immunoglobulin homology <IMM1>
F:133-191/Domain: immunoglobulin homology <IMM2>
F:153-156/Region: heparin binding #status predicted
F:161-165/Region: heparin binding #status predicted
F:228-290/Domain: immunoglobulin homology <IMM3>
F:263-272/Region: NCAM binding #status predicted
F:323-388/Domain: immunoglobulin homology <IMM4>
F:420-482/Domain: immunoglobulin homology <IMM5>
F:519-596/Domain: fibronectin type III repeat homology <FN3A>
F:653-685/Domain: fibronectin type III repeat homology <FN3B>
F:721-96,133-189,235-288,330-386,427-480/Dissulfide bonds: #status predicted
F:442,316,348,424,450,479/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match	5.4%	Score 125.5	DB 1	Length 725
Best Local Similarity	20.9%	Pred. No. 1.3		
Matches 95	Conservative 64	Mismatches 187	Indels 109	Gaps 21

```

0Y 3 FG-VPRHLLVLQALLPATOQKVVJGKKDVELLTCTASOKKSIOPFH-KSNNOJK 60
Db 196 RGEINFKIOIYIVNPTVQAFOSIVNATANIGOSVTLVCDADGPEPTMSWKOEPIE 255
0Y 61 ILNGOSFELTGPSTKANDRAD--SRSLMDQGNFPLIKNLKIEDSDTYICEVE---DQ 114
Db 256 -----NEEDBERGSSVSDSE--VTIRNVKDNDBAEYVCIANENKAGEQ 297
0Y 115 KEVOVLVFG---LTSNPTHLOGOSILTLTLESBPSSBPVOQR-----SPRKNIQ 164
Db 298 DASHLKVFAPKRTIYVENQJAMELEOVTLTCEASGDEPISITWRTSFRNISSEBODID 357
0Y 165 G3-----KTLVSQLELODSSGTWTCTVQONQKVEFKIDIVIAFOKASSI----- 210
Db 358 GHMVYRSHARVSSLTLSKSIQYRDABEYMTASNTIGQDSQSID--LEQYAPKLOGPVA 414
0Y 211 VYKKEGEVE---FSFPLAFTVEKLTGSGELMWAERASSKSWITFDLKN-KEVSVK 264
Db 415 VYTWEGNOVNITCEVEAFYPSA-TIS-----WFRDQGLLPSNSNIKIYVTPSASYL 465
0Y 265 RYTODPKIQMGK-----KLPLHLTLPOA-----LPQYAGSGNLTLLAEK 304
Db 466 EYTPDSENDFGVYNTACTAVNRIGQESLEFLVQADTPSPSSIDRVEYSSTAOYQFBEPPEA 525
0Y 305 TGKTLHQEVNLYVMRATOLOKUNLTCEVWGPTSPDKMLSLTLENKEAKVYSKREKRVWVLANPE 364
Db 526 TG-----GVPLIKYKAEKMSJGEBSSWHT-----WYDAKEANMEGIYV-----IMGLKDE 570
0Y 365 AGMOCCLSDSGOVL-----LESNIKVLPTMSTP 393
Db 571 TTYSDRLAALNGKGGEIMQPEBSSTQVPPPELSAP 605

```

RESULT 36
A54100
tumor suppressor protein DCC precursor - human
N.Alternate names: colorectal cancer suppressor DCC
C.Species: Homo sapiens (man)
C.Date: 02-Aug-1994 #sequence_revision 02-Aug-1994 #text_change 05-Nov-1999
C.Accession: A54100; A40098
R.Herrick, L.; Cho, K.R.; Fearon, E.R.; Wu, T.C.; Kinzler, K.W.; Vogelstein, B.
Genes Dev. 8, 1174-1183, 1994
A.Title: The DCC gene product in cellular differentiation and colorectal tumorigenesis.
A.Reference number: A54100; MUID:95011532; PMID:7926732
A.Accession: A54100
A.Molecule type: mRNA
A.Residues: 1-1447 <HED>
A.Cross-references: EMBL:X76132; NID:g453209; PIDN:CAAS3735.1; PID:g453210
R.Fearon, E.R.; Cho, K.R.; Nigro, J.M.; Kern, S.E.; Simons, J.W.; Ruppert, J.M.; Hamilton

A:Title: Identification of a chromosome 18q gene that is altered in colorectal cancers.

A:Reference number: A40098; MUID:90100559; PMID:2294591

A:Accession: A40098

A:Molecule type: mRNA

A:Residues: 1-750 <FEA>

A:Cross-references: GB:M32292; NID:G181492; PIDN:AAA3751.1; PID:G181493

C:Genetics:

A:Gene: GDB:DC

A:Cross-references: GDB:119818; OMIM:120470

A:Map position: 18q21.1-18q21.1

C:Keywords: transmembrane protein; tumor suppressor

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-144/Product: tumor suppressor protein DCC #status predicted <MAT>

Query Match 5.4%; Score 125.5; DB 2; Length 1447;

Best Local Similarity 19.3%; Pred. No. 3.3; Mismatches 129; Indels 147; Gaps 15;

Matches 77; Conservative 46; Mismatches 129; Indels 147; Gaps 15;

Qy 34 GDTVELTCTASQKSIQPHWKNSNQIKILGNQSFITKPSKLNDRADSRSLMDQGNFP 93

Db 154 GDTVLKCEVIGEPHPTIHMQNQ-----DLTRIP-----GDSRVVLPSSG--A 196

Qy 94 LIINKLKIEDSDTYICEVD-----QKEVOLI-----VGLTANSDTHLQGS 138

Db 197 LOISRLQPGDIGYRCASRNPAASRTGNEAEVRLISDPGLRHOLYFLORPENVVAIEGKD 256

Qy 139 LTL--TLESPPGS-----SPSVQCRSPRGKNIQGGKTLVSQLELDDSGTWTCTVQN 189

Db 257 AVLECCVGYPPPSFTWLRGEEVITQLRSK--YSLGGSNLISNTDDSGMYTCVTVYK 315

Qy 190 QKRVFKIDIVLA---FQKASSIVYKKEGQVEFSPFLAFTVEKLTGSGELMWOAERAS 246

Db 316 NENISASAEITLVLPWFLNHPNLSLYAESMDIE-----ECTYSG-----KPV 359

Qy 247 SSKSKITTDLKNKEYSVKRVITQDPKIQMGKKLPLHLTLPQAYAGSGLTLAEAKTG 306

Db 360 PTVNM---MKNGDV-----VIPSDFYFQIVGSGNIRI----- 387

Qy 307 KLHGEVNLVVRATQLOKNLCEVWGPSPKMLSLKLENKAQVSKKEKVVWVLNPEAG 366

Db 388 -----LGVYKSDGE 396

Qy 367 MMQCLSDSGOVLLESNIKVLPTWSTPVHPRASALPAP 405

Db 397 FYQCVAEENAGNAQSAQLIVP--KPAIPSSSVLPSPAP 432

RESULT 37

108851

Down syndrome cell adhesion protein 1 - human (fragment)

N:Alternate names: Down syndrome cell adhesion molecule

C:Species: Homo sapiens (man)

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999

C:Accession: T08851

R:Yamakawa, K.; Hsu, Y.K.; Haendel, M.A.; Hubert, R.; Chen, X.N.; Lyons, G.E.; Korenberg

submitted to the EMBL Data Library, September 1997

A:Description: DSCAM: A novel member of the immunoglobulin superfamily maps in a down sy

A:Reference number: Z16495

A:Accession: T08851

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1896 <YML>

A:Cross-references: EMBL:AF023449; NID:G3169765; PID:G3169766

A:Experimental source: brain; developmental stage: 14 weeks; fetal

C:Genetics:

A:Gene: DSCAM

A:Map position: 21q22

A>Note: derived from alternately-spliced mRNA

C:Function:

A:Description: involved in nervous system development

C:Keywords: alternative splicing

Best Local Similarity 18.6%; Pred. No. 5.1; Mismatches 82; Conservative 55; Mismatches 144; Indels 160; Gaps 17;

Matches 82; Conservative 55; Mismatches 144; Indels 160; Gaps 17;

Qy 20 PATQGNKVVLGKKGDTVELTCTASQKSIQPHW-KNSNQIKILGNQSFITKPSKLNDR 78

Db 490 PASIRPKNITTAJAGRDYTHCRVIGYPPYSIKRYKNSNLLPFNHRQVAFENNGTLLKSD 549

Qy 79 RADSRSLMDQGNPFLIKNLKIEDSDTYICEVDQKEEVQLVFGLTANSDTHLQGS 138

Db 550 VQ-----KEYDGEYTCNVLPQ-----LSTQS 574

Qy 139 LTLTLESPP-----GSSPSVQCRSPRG-----KNI 163

Db 575 VHTVKKVPFIQPEPFPSIIGQRFVIPCVAAGDLPITITWQKGRPIGSLGVTIDNI 634

Qy 164 QGGKTLVSQLELDDSGTWTCTVQLQNKKEFKIDIVLA-----FQKASS 209

Db 635 DFTSLRLISNLSIMHNGNYTCIARNEAAVEHQSOLIVRPKRVVQPRDQGIYGA VI 694

Qy 210 IYVKEGQVEFSPFLAFTVEKLT--GSGELMWOAERASSKSWITFDLKNKEYSVK----- 264

Db 695 LNCSAEG-----YVPTIYWKESKAGVPOFO--PILNKRIOV--LSNGSLIKHIVE 744

Qy 265 -----RYTODPKIQMGKKLPLHLTLPQALPQYAGSGLTLAEAKTGKLHGEVNLV 316

Db 745 EDGQYLVCKVSNIDVADVSKMYLTIVKIPAMITSYF--NTTLATQGO----- 789

Qy 317 MRATQLOKNLCEVWGPSTFKMLSLKLENKAQVSKKEKPV--W-----VLPENAGMW 366

Db 790 -----KKEMSCTAHG-----EKPIIVWEKEDRIIPENARY 821

Qy 369 QCLSDSGOVLLESNIKVLPT 389

Db 822 LVSTKEVGEVIT-STLQILPT 841

RESULT 38

520901

ctln - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999

C:Accession: S20901; I46520

R:Label, S.; Barlow, D.P.; Gautel, M.; Lakey, A.; Trinick, J.

EMBO J. 11, 1711-1716, 1992

A:Title: Towards a molecular understanding of titin.

A:Reference number: S20897; MUID:92258380; PMID:1582406

A:Accession: S20901

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-6805 <LAB>

A:Cross-references: EMBL:X64696

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992

R:Label, S.; Barlow, D.P.; Gautel, M.; Gibson, T.; Holt, J.; Heien, C.L.; Francke, U.; I

Nature 345, 273-276, 1990

A:Title: A regular pattern of two types of 100-residue motif in the sequence of titin.

A:Reference number: I46520; MUID:90238553; PMID:2129545

A:Accession: I46520

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 4235-5250 <LA2>

A:Cross-references: EMBL:X17329; NID:G1756; PIDN:CAA35207.1; PID:G930251

C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; prot

C:Keywords: muscle

Query Match 5.4%; Score 125; DB 2; Length 6805;

Best Local Similarity 18.6%; Pred. No. 28; Mismatches 11; Conservative 81; Mismatches 180; Indels 224; Gaps 26;

Qy 11 LTVLQALPLPA--TGG--NNVVLGKKGDTVELTCTASQKSIQPHWKNSQIKILGNQGS 67

Db 2916 VIVKEQTLPELDLRGRYQKLVATAKGNIKIVEIPVLGRKXPVTWKKGQVLLKQTRVN 2975

Qy 68 FLTKGSPKLNDRADSRSLMDQGNPFLIKNLKIEDSDTYICEVD-----QKEEV 118


```

Db 2976 VENTATSTILINISSEVR--DSGPRFLAKNI VGEVGVITTIQVNDIGPPTGPIKPEV 3033
Qy 119 QLVYFGLTANSPDTHLLQGOSLTLTLESPPGSSPVQCSSPRKNGIKQGTLS--VSQLEL 176
Db 3034 -----SSD-----FVTFSEWEPEN-----DGVVPSISNVIENRQ 3062
Qy 177 QDSGTV---TCTVLONOKK-----VEFKIDI-----VLAFOKXSSI----- 210
Db 3063 TDSITWELATVITVITTKATRLTGTVEYQFRVKAQNRVGVPGITTSASIVANPFVKVG 3122
Qy 211 -----VYKKEGQVEFSFPLA-----FTVEKLTGSGELMWQAEBA----- 245
Db 3123 PGCTPQVTAVTKDSMTISNHEPLSDGSGPILOGHYERKERNDILMQTVSKALVPENIFKS 3182
Qy 246 SSSKSWITFDL-----KXKEYSVKRVTDQPKLQMGKPLHL-----TLPOALP 289
Db 3183 SGLTDGIAYEFVIAENNAKSGKSPKSPVLDLPIDPPGKPIPLNITRHTVTLKMAKP 3242
Qy 280 QYAGSGNLTLMLAK-----TGKLNQ-----EVNLVYMRATQ-----LQKMLTCEV 330
Db 3243 ETGGSPKITSYIVEXRDLPNQWMLKANSNILENEFTVSGLTEDAAVEFRVIAKNAAGAI 3302
Qy 331 WGPV-----SPKMLSLKLENK-----EAKYSKEEKP----- 357
Db 3303 SPSPSPDAITCRDVEAPRILVDVRFKDTVLKGAFAKLEADVSGRPPTMEVTKDGK 3362
Qy 358 -----VWVLNPEAGMOCCLSDSQVLLS-----NIKVLPTW 390
Db 3363 ELECTGKLEIKIADSFYLIINDSSR-----RDSGAVILTATDPGFAKHFNVLVLR- 3416
Qy 391 STPVHPRASALPAPRTGSLPPTQASALPDPAPASALPAAVLSFLGLGLGVA 446
Db 3417 -----PGPEEG-----PLAVSEVTSEKCVLSWLPDDGSA 3447

RESULT 39
LJMSNL
neural cell adhesion molecule 1 precursor, long domain splice form - mouse
N:Alternate names: NCAM-180
N:Contains: neural cell adhesion molecule, short domain splice form (NCAM-140)
C:Species: Mus musculus (house mouse)
C:Date: 31-Mar-1993 #sequence revision 31-Mar-1993 #text change 31-Dec-2000
C:Accession: A29673; S00844; S00384; A28281; A44290; S00383
R:Bartheleis, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chalk, J.C.; Hirsch, M.R.; Fontec
EMBO J. 6, 907-914, 1987
A>Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,000
A:Reference number: A29673; MUID:87246524; PMID:359563
A:Accession: A29673
A:Molecule type: mRNA
A:Residues: 1-548, 'T', 550-571, 'T', 573-574, 'D', 576-588, 'MQPS', 593, 'S', 595-599, 'P', 601, 'L'
A:Cross-references: EMBL:X00051; NID:953342; PIDN:CAA68263.1; PID:953343
R:Santoni, M.J.; Bartheleis, D.; Barbas, J.A.; Hirsch, M.R.; Steinmetz, M.; Goridis, C.; W
Nucleic Acids Res. 15, 8621-8641, 1987
A>Title: Analysis of cDNA clones that code for the transmembrane forms of the mouse neur
A:Reference number: S00844; MUID:88067687; PMID:3684567
A:Accession: S00844
A:Molecule type: mRNA
A:Residues: 529-809, 1077-1115 <SN>
A:Cross-references: EMBL:X06328; NID:953322; PIDN:CAA29641.1; PID:9817984
R:Barbas, J.A.; Chalk, J.C.; Steinmetz, M.; Goridis, C.
EMBO J. 7, 625-632, 1988
A>Title: Differential splicing and alternative polyadenylation generates distinct NCAM tr
A:Reference number: S00382; MUID:88283628; PMID:3396534
A:Accession: S00384
A:Molecule type: DNA
A:Residues: 642-1115 <BAR>
A:Cross-references: EMBL:X07195
R:Bartheleis, D.; Vopper, G.; Wille, W.
Nucleic Acids Res. 16, 4217-4225, 1988
A>Title: NCAM-180, the large isoform of the neural cell adhesion molecule of the mouse,
A:Reference number: A28281; MUID:88247737; PMID:2454455
A:Accession: A28281

```

```

A:Molecule type: mRNA
A:Residues: 804-1081 <BA3>
A:Cross-references: EMBL:X07244; NID:953321; PIDN:CAA30230.1; PID:g929720
R:Rougon, G.; Marshak, D.R.
J. Biol. Chem. 261, 3396-3401, 1986
A>Title: Structural and immunological characterization of the amino-terminal domain of m
A:Reference number: A44290; MUID:86140120; PMID:3512556
A:Accession: A44290
A:Molecule type: protein
A:Residues: 20-36 <ROD>
C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
C:Comment: Several forms of NCAM are produced by alternative splicing. See also PIR:LJMS
C:Genetics:
A:Gene: NCAM
A:Map position: 9
A:Introns: 643/3; 701/1; 770/2; 809/2; 1076/2
C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C:Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; st
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1115/Product: neural cell adhesion molecule, long domain splice form #status experi
F:20-809/1077-1115/Product: neural cell adhesion molecule, short domain splice form #sta
F:20-711/Domain: extracellular #status predicted <EXT>
F:34-98/Domain: immunoglobulin homology <IMM1>
F:132-191/Domain: immunoglobulin homology <IMM2>
F:152-156/Region: heparin binding #status predicted
F:161-165/Region: heparin binding #status predicted
F:228-290/Domain: immunoglobulin homology <IMM3>
F:262-272/Region: NCAM binding #status predicted
F:323-388/Domain: immunoglobulin homology <IMM4>
F:420-482/Domain: immunoglobulin homology <IMM5>
F:519-596/Domain: fibronectin type III repeat homology <FN3A>
F:625-665/Domain: fibronectin type III repeat homology <FN3B>
F:712-729/Domain: transmembrane #status predicted <TM>
F:730-1115/Domain: intracellular #status predicted <INT>
F:41-96, 139-189, 235-288, 330-386, 427-480/Disulfide bonds: #status predicted
F:222, 316, 348, 424, 450, 479/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.3%; Score 124; DB 1; Length 1115;
Best local similarity 21.9%; Pred. No. 2.9;
Matches 68; Conservative 47; Mismatches 128; Indels 68; Gaps 15;

Qy 3 RG-VFPHILVLDLALPAATGKVKVGLGKGTVELTCTASQKSIQFM-KNSNQIK 60
Db 196 RGEINFDIOIVVAVPTVQARQSIIVANATLGGSVTLVCDADGPPTMTGDCGEIE 255
Qy 61 ILNGSGFLTYGPKSLNDRAD-SRSLMDQGNFLIIXLKIEDSDTYIEVE---DQ 114
Db 256 -----NEEDERSRSVSDDSE--VITRNDKXNDEAYVCIAENKAGEQ 297
Qy 115 KEVQLLVFG---LTANSDFHLQGGSLTLTLESPPGSSPVQCR-----SPRGNKIQ 164
Db 298 DASIHLLKVFAPKKTYYENQJAMELEBEVTLTCAAGDPISITWRSTRNISSEBQDL 357
Qy 165 GG-----KTSVSGLELQDSGTWTCYVLQONQKVEFKIDIVLAFOKASSI----- 210
Db 358 GHMVRSHARVSSLTLSIQYRDAGEYCTASNTIGDQSGSID---LEFOYAPKLQGPVA 414
Qy 211 VYKKEGQVE-----FSFPLAFTVEKLTGSGELMWQAEARASSSKSWITFDLKN-KEYSVK 264
Db 415 VYTWEGNQVNITCEFAVPASA-TIS-----WFRDQQLPSNYSNIKIYNTPASAYL 465
Qy 265 RVTQDPKLGKMG 275
Db 466 EVTPDSENDFG 476

RESULT 40
A64157
hypothetical protein HI0696 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence revision 18-Aug-1995 #text change 29-Sep-1999
C:Accession: A64157
R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A

```



```

Db      58 EG-----KAPQRLLYVDVSTARDVLESGSPGKYTHTRRMRMILRLQNLINDSGVYYC 113
Qy      110 EVEDQKEVQLLVEFLTANSDTHLLQGSGSLT---LESPSSSSVQCRSPRGKNIOGG 166
Db      114 ATWDRPRLKLT--FG-----SGTTLVTDQQLADVSPKPTIFLP----- 151
Qy      167 KTLSSQLELQDGTGWTCTVLQNKKEFKIDIVLAFC--KASSIVKKEGE----- 217
Db      152 ---SIAEKLQAKAGTYLLCLL-----EKFPPIIKIHMOEKSSNTILQSGSBNMTKMTDT 202
Qy      218 QVESFPPLAFVTEKLTGSGELMWOAERASSKSMITFDLKNKESVKEVETDPPKLQMGKK 277
Db      203 YMKFSM-LTVPRESLDKEHRC--IVRHNNKNGIDQELIPEPITDVTDTYDPPKYNKSD 258
Qy      278 LPLHLTQALPQYAGSGNLTALPAKTKGLHQEVNLYVMRATQQLKLTLC 328
Db      259 ANDVITM-DPKMWSKANDTLLQLTNTSAVYTYLLLLKSVVFAITTC 308

RESULT 43
ITCNLU
neural cell adhesion molecule long domain form precursor - chicken
N:Alternate names: NCM-180
N:Contains: neural cell adhesion molecule, short domain form (NCAM-140)
C:Species: Gallus gallus (chicken)
C>Date: 31-Mar-1993 #sequence revision 31-Mar-1993 #text change 22-Jun-1999
C:Accession: A43613; B43613; A25435; B25435; S36950; A46550; A60852; S29668
R:Cunningham, B.A.; Hemperly, J.C.; Murray, B.A.; Prediger, E.A.; Brackenbury, R.; Edelman
Science 236, 799-806, 1987
A>Title: Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell sur
A:Reference number: A43613; MUID:87206190; PMID:576159
A:Accession: A43613
A:Molecule type: mRNA
A:Residues: 1-175 <CU>
A:Cross-references: GB:M15860
A:Accession: B43613
A:Molecule type: protein
A:Residues: 20-44;120-127;202-221;320-342;399-415;640-659;822-828 <CUN>
A>Note: Asn-222 probably binds carbohydrate; Asn-226 probably does not
R:Hemperly, J.C.; Murray, B.A.; Edelman, G.M.; Cunningham, B.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 3037-3041, 1986
A>Title: Sequence of a cDNA clone encoding the polysialic acid-rich and cytoplasmic doma
A:Reference number: A25435; MUID:86506089; PMID:3458261
A:Accession: A25435
A:Molecule type: mRNA
A:Residues: 128-1091 <HEM>
A:Cross-references: GB:M13210
A:Accession: B25435
A:Molecule type: protein
A:Residues: 128-140;222-240;428-439;611-631;744-760;763-781;1080-1084 <HE2>
R:Murray, B.A.; Owens, G.C.; Prediger, E.A.; Crossin, K.L.; Cunningham, B.A.; Edelman, G
J. Cell Biol. 103, 1431-1439, 1986
A>Title: Cell surface modulation of the neural cell adhesion molecule resulting from alt
A:Reference number: A46550; MUID:87033934; PMID:3771645
A:Accession: A46550
A:Molecule type: DNA
A:Residues: 810-1070 <MUR>
A:Cross-references: GB:X04479
R:Saenger, M.; Covault, J.
submitted to the EMBL Data Library, February 1993
A:Reference number: S36950
A:Accession: S36950
A:Molecule type: DNA
A:Residues: 1-17 <SAS>
A:Cross-references: EMBL:X70342; NID:G417631; PIDN:CAA49807.1; PID:G417632
R:Colwell, G.; Li, B.; Forrest, D.; Brackenbury, R.
Genomics 14, 875-882, 1992
A>Title: Conserved regulatory elements in the promoter region of the N-CAM gene.
A:Reference number: A44369; MUID:93122797; PMID:1478668
A:Accession: A44369
A:Molecule type: DNA
A:Residues: 1-17 <COF>
A:Cross-references: EMBL:Z12128; NID:G63653; PIDN:CAA78113.1; PID:G63654

```

```

A:Experimental source: White Leghorn
R:Col, G.J.; Loewy, A.; Cross, N.V.; Akesson, R.; Glaser, L.
J. Cell Biol. 103, 1739-1744, 1986
A>Title: Topographic localization of the heparin-binding domain of the neural cell adhes
A:Reference number: A60852; MUID:87057627; PMID:2430978
A:Accession: A60852
A:Molecule type: protein
A:Residues: 20-29 <COL>
R: Rao, Y.; Wu, X.F.; Gartezy, J.; Rutishauser, U.; Sin, C.H.
J. Cell Biol. 118, 937-949, 1992
A>Title: Identification of a peptide sequence involved in homophilic binding in the neur
A:Reference number: A43280; MUID:92363934; PMID:1380002
A:Accession: A43280
A:Contents: annotation: homophilic binding region
C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
C:Superfamily: Various forms of NCAM are produced by alternative splicing.
C:Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; si
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1091/Product: neural cell adhesion molecule, long domain form #status experimental
F:20-809;1071-1091/Product: neural cell adhesion molecule, short domain form #status exp
F:20-711/Domain: extracellular #status predicted <EXT>
F:34-98/Domain: immunoglobulin homology <IMM1>
F:132-191/Domain: immunoglobulin homology <IMM2>
F:152-156/Region: heparin binding #status predicted
F:161-165/Region: heparin binding #status predicted
F:228-289/Domain: immunoglobulin homology <IMM3>
F:262-271/Region: NCAM binding #status experimental
F:322-387/Domain: immunoglobulin homology <IMM4>
F:419-481/Domain: immunoglobulin homology <IMM5>
F:518-595/Domain: fibronectin type III repeat homology <FN3A>
F:624-665/Domain: fibronectin type III repeat homology <FN3B>
F:712-729/Domain: transmembrane #status predicted <TM>
F:730-1091/Domain: intracellular #status predicted <INT>
F:41-96;139-189;235-287;329-385;426-479/Dissulfide bonds: #status predicted
F:222/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:315;347;423;449;478/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.3%; Score 123.5; DB 1; Length 1091;
Best Local Similarity 20.2%; Pred. No. 3;
Matches 85; Conservative 51; Mismatches 157; Indels 127; Gaps 16;

Qy      32 KKQDTVELTTPAQSQKSIQFMKNSNQIKINGSGFLTKGPKSLNBRADSRSLMDGN 91
Db      130 KEGDAVIVCDVSSLPPTITWKKGRDVLKQVRIV-----LSN 171
Qy      92 FPLIINKLIEDSDTYICE-----VEDQKEVOLV-----FGLTANSDTHL 133
Db      172 NYLQIRGIKKTDEGTVCGRILARGEINFDIOVIVNPPSVARQSTWMTATN----- 226
Qy      134 LQGSGLTILTESPPSSPSVQCRSPRGKNIO-----GKTLSSQLELQDSGT 181
Db      227 -LSGSVTLACDADPPEPTTWM--TKDGEPIQEDNEKRYSPNYGSEILIKVAKSDPAE 284
Qy      182 WTCVTLQNKKEVKRIDIIVLAFOKASSIVYKKEGGEVSEFPPLAFVTEKLTGSGE--- 237
Db      285 YICIAENKAGQDATHLKVFAKPK---ITYVEKTMLELEDQITLICE---ASGDPIPS 338
Qy      238 LMQAERASSSKSMITFDLKNKESVVRVTDPPKLQMGKKLPLHLTQALPQYAGSGNT 297
Db      339 ITWKTSTRNISNEKKTLDGRIVRSHARVSS-----LTLEKEI--QYTDAGEY 383
Qy      298 TLAEAKTKGLHQEVNLYVMRATQ-----QKULTCVWGPSTSKMLSLKLE 345
Db      384 VCTASNTIGDSQAMYLEVOYAPPLQGPVAVYTWEGNQVITCEVF--AYPSAIVS---- 437
Qy      346 NKEAKVSKREKPVVVLNPEAGMQLSDSGQVLTLESNIKVLPTWSTPVPFRASALPAP 405
Db      438 -----WFRD-----GQLPSSNYSNIKIYNT---PSASYLEVTP 468

RESULT 44
A38096
perlecan precursor - human

```

N.Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate proteoglycan
C.Species: Homo sapiens (man)
C.Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 05-Nov-1999
A.Accession: A38096; S19256; S77946; A41059; A40306; B33625; A41736
R.Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J.Biol.Chem. 267, 8544-8557, 1992
A.Title: Primary structure of the human heparan sulfate proteoglycan from basement mem-
brane, laminin, neural cell adhesion molecules, and epidermal growth factor.
A.Reference number: A38096; MUID:92235084; PMID:1569102
A.Accession: A38096
A.Molecule type: mRNA
A.Residues: 1-4391 <MUR>
A.Cross-references: GB:M5289; NID:g184426; PIDN:AAA52700.1; PID:g184427
R.Kalunki, P.; Trygvaason, K.
J.Cell Biol. 116, 559-571, 1992
A>Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-KD pro-
tein with multiple extracellular domains and an intracellular domain.
A.Reference number: A41736; MUID:92112994; PMID:1730768
A.Accession: S19256
A.Molecule type: mRNA
A.Residues: 1-57, D, '59-434, 'A, '436, 'FL, '438-449, 'Q, '451-502, 'A, '503-792, 'K, '794-908, 'R,
'71-2979, 'H, '2981-2994, 'G, '2996-3167, 'T, '3169-3240, 'R, '3242-3426, 'R, '3428-3631, 'Q, '3633-3634
A.Cross-references: EMBL:X62515; NID:g29469; PIDN:CMA43373.1; PID:g29470
R.Kalunki, P.; Eddy, R.L.; Byers, M.G.; Kesilae, M.; Shows, T.B.; Trygvaason, K.
Genomics 11, 369-396, 1991
A>Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the g-
ene.
A.Reference number: A41059; MUID:92120660; PMID:1685141
A.Accession: A41059
A.Molecule type: mRNA
A.Residues: 'RT, '892-908, 'R, '910-1101, 'L, '1103-1132, 'L, '1134-1221, 'L, '1223-1397 <KA>
A.Cross-references: GB:S76436; NID:g243370; PIDN:AA821121.1; PID:g243371
R.Dodge, G.R.; Kovalevsky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo, R.V.
Genomics 10, 673-680, 1991
A>Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellular localization, and synthesis.
A.Reference number: A40306; MUID:91365376; PMID:1679745
A.Accession: A40306
A.Molecule type: mRNA
A.Residues: 1018-1405, 'G, '1407-1409, 'G, '1411-1465 <DD>
A.Cross-references: GB:M4283; NID:g184424; PIDN:AAA5269.1; PID:g184425
R.Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den Berghe, H.
Cell Biol. 109, 3199-3211, 1989
A>Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclonal antibodies.
A.Reference number: A33625; MUID:90078352; PMID:2687294
A.Accession: B33625
A.Molecule type: protein
A.Residues: 1379-1384, 'X, '1386-1388, 'X, '1390-1398 <HE2>
A.Accession: A33625
A.Molecule type: protein
A.Residues: 2166-2171, 'X, '2173-2175, 'X, '2177-2185 <HE3>
A.Note: peptide potentially matches four different regions of sequence shown
C.Genetics:
A.Gene: GDB:HSPG2
A.Cross-references: GDB:126372; OMIM:142461
A.Map position: 1p36.1-p36.1
C.Summary: LDL receptor ligand-binding repeat homology; EGF homology; laminin G repeats; heparan sulfate; transmembrane protein
C.Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembrane protein
F.II-21/Domains: signal sequence #status predicted <SIG>
F.II-22-4391/Product: perlecan #status predicted <MAT>
F.II-22-193/Domains: I <DOM1>
F.II-22-193/Domains: II <DOM2>
F.II-22-193/Domains: III <DOM3>
F.II-22-193/Domains: IV <DOM4>
F.II-22-193/Domains: V <DOM5>
F.II-22-193/Domains: VI <DOM6>
F.II-22-193/Domains: VII <DOM7>
F.II-22-193/Domains: VIII <DOM8>
F.II-22-193/Domains: IX <DOM9>
F.II-22-193/Domains: X <DOM10>
F.II-22-193/Domains: XI <DOM11>
F.II-22-193/Domains: XII <DOM12>
F.II-22-193/Domains: XIII <DOM13>
F.II-22-193/Domains: XIV <DOM14>
F.II-22-193/Domains: XV <DOM15>
F.II-22-193/Domains: XVI <DOM16>
F.II-22-193/Domains: XVII <DOM17>
F.II-22-193/Domains: XVIII <DOM18>
F.II-22-193/Domains: XIX <DOM19>
F.II-22-193/Domains: XX <DOM20>
F.II-22-193/Domains: XXI <DOM21>
F.II-22-193/Domains: XXII <DOM22>
F.II-22-193/Domains: XXIII <DOM23>
F.II-22-193/Domains: XXIV <DOM24>
F.II-22-193/Domains: XXV <DOM25>
F.II-22-193/Domains: XXVI <DOM26>
F.II-22-193/Domains: XXVII <DOM27>
F.II-22-193/Domains: XXVIII <DOM28>
F.II-22-193/Domains: XXIX <DOM29>
F.II-22-193/Domains: XXX <DOM30>

Query Match 5.3%; Score 122.5; DB 2; Length 4391;
Beet Local Similarity 22.5%; Pred. No. 23; Mismatches 142; Indels 147; Gaps 21;
Matches 98; Conservative 49;

F:533-1676/Domain: III <DOM3>
F:1159-1206/Domain: laminin-type EGF-like homology <LEG>
F:1553-1610/Domain: laminin-type EGF-like homology <EG7>
F:1613-1668/Domain: laminin-type EGF-like homology <LEG8>
F:1677-1686/Domain: IV <DOM4>
F:2007-2034/Domain: transmembrane #status predicted <TRM>
F:3687-4391/Domain: V <DOM5>
F:3845-3880/Domain: EGF homology <EGF1>
F:3888-3921/Domain: EGF homology <EGF>
F:3953-4106/Domain: laminin G repeat homology <LG3>
F:4147-4175/Domain: EGF homology <EGF2>
F:4149-4151/Region: motor neuron attachment (L-R-E) motif
F:4299-4301/Region: motor neuron attachment (L-R-E) motif
F:655-71,76/Binding site: Heparin sulfate (Ser) (covalent) #status predicted (asn) (covalent)
F:878,554,1755,2121,3072,3105,3279,3780,3835,4068/Binding site: carbohydrate
F:2995,3933,4179/Binding site: Chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 5.3%; Score 122.5; DB 2; Length 4391;
Beet Local Similarity 22.5%; Pred. No. 23; Mismatches 142; Indels 147; Gaps 21;
Matches 98; Conservative 49;

12 LVLQALPAAATGKGVVVGKADTVETLCTASQKSIQFHWKNSQIKILGNGSEFLTK 71
3291 IILHESPPVATTVENHSAVQAGETVQLQCLAHGTPLRTGQ-----SR 3334
72 GPSKLNDRADSRSLMDQGNPLIINKLIEDSTYICEVEDQ---KEEVQLIVFGLTA 127
3335 VGSSILPGRATATNEL-----LHFERAAPEDSGRYRCVTNKVGSAAFAQLLVQC-- 3384
128 NSDTHLLDQGSLLTLSPSPGSSPVQCSPRKNIQGGKTSVLS---QLELDQSGWTC 184
3385 -----PPGSLPAT-----SIPAGSPVQVPTQLETSGIGA-- 3415
185 TVLQNKQKFEFKIDIVLAFQKASSIVYKKEGEQVFFSLATVEKLTGSGELMWAER 244
3416 -----SVEFH---CAVPSDQGTQLRMRKEGQL-----PRGHSVQ---DGVLRIGNLD 3457
245 ASSSKSWITFDIKNKEVSKVRYTQDP--KLQWKGKPLHLTLRQALPOYAGSGNLTAL 302
3458 QSCQGTLYC-----QAHGFWGKAQASQALVI-----QALP-----SVLINIRTS 3496
303 AKTGKLNQEVNLVWVRATQLOKNLTCGEVWGPPISPKMLSLKLNKKAQVSKREPVWLN 362
3497 VQT-----VVGHAAVEF---CLALGDRPKRYTWS-----KVGHNLPRGIQVS 3536
363 -----PEAGWMOCLSDS-----GOVLLSEINIKVLPMTSTPVHPRASALPAPT 406
3537 GGAVRIAHVELADAGQYRCTATINNAAGTQSHVLL--LVQALPQISMGEVRY-----PA 3588
407 GSALPDPQTASALPDP 422
3589 GSAAVFPCTIASGYTPR 3604

RESULT 45
PNO568
connectin 3B - chicken (fragment)
N:Alternate names: Cn3B protein
C:Species: Gallus gallus (chicken)
C:Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 17-Mar-1999
C:Accession: PNO568
R:Matuyama, K.; Endo, T.; Kume, H.; Kawamura, Y.; Kanzawa, N.; Nakauchi, Y.; Kimura, S.;
Biochem. Biophys. Res. Commun. 194, 1288-1291, 1993
A:Title: A novel domain sequence of connectin localized at the 1 band of skeletal muscle
A:Reference number: PNO568; MUID:93356802; PMID:8552787
A:Accession: PNO568
A:Molecule type: mRNA
A:Residues: 1-1323 <MAR>
A:Cross-references: DBJ:P16541; NID:g391629; PID:d1004495; PID:g391630
A:Experimental source: skeletal muscle
C:Comment: This protein encoding-like single molecule spans from the Z line to the M line

Query Match 5.2%; Score 122; DB 2; Length 1323;

Best Local Similarity 18.7%; Pred. No. 4.9;
Matches 100; Conservative 59; Mismatches 153; Indels 222; Gaps 22.

[illegible]

RESULT 46
JC4776
limbic-system-associated membrane protein precursor - human
C|Species: Homo sapiens (man)
C|Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 19-May-2000
C|Accession: J04776
R|Pimenta, A.F.; Fischer, I.; Levitt, P.
Gene 170, 189-195, 1996
A|Title: cDNA cloning and structural analysis of the human limbic-system-associated mem
A|Reference number: J04776; MUID:96235133; PMID:8666243
A|Accession: J04776
A|Molecule type: mRNA
A|Residues: 1-338 <PIM>
A|Cross-references: GB:U41901; NID:G1276898; PIDN:AAC50569.1; PID:G1276899
A|Experimental source: brain
C|Comment: This is a neuronal surface glycoprotein distributed in cortical and subcortical
C|Genetic8:
A|Gene: Lamp
C|Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
C|Keywords: brain; glycoprotein; membrane protein; phosphoprotein
F|1-7|Domain: signal sequence #status predicted <SIG>
F|333-338|Region: hydrophobic
F|40,66,136,149,219,287,300,315|Binding site: carbohydrate (asn) (covalent) #status pred
F|42,115,142,164,171,220,231|Binding site: phosphate (Thr) (covalent) #status predicted
F|95,129,204,226,310|Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 5.2%; Score 121.5; DB 2; Length 338;

Best Local Similarity 20.7%; Pred. No. 0.85;
Matches 86; Conservative 63; Mismatches 123; Indels 143; Gaps 22

OY 10 LLLVLQLALLPAA-----TQGNKVVLGKKGPTELVTCTASQSKS IQFHMKNSNQIKI 61

DB 14 LVLLRLCLLPGLPVRSVDFNREGTDNI TVRGGTAIRKCVLEDKNS -KYAMLNRSGIIIF 72

```

Qy 62 LGHQSGFLTKGSGSKLNDRA--DSKSLMDQGNFPLITLKNLKBESDPTYLICEVDQGE--- 116
Db 73 AGHD-----KMSIDPFRYLEKGRHSL-----EYSLRIOKQDVYDEGYSYTSVOTQHPKPT 121
Qy 117 -EYQLLVFG-----LTRANSPTHLLOQSLTLTLSPSPGSSPSVQGR---SPGKGNIOGSKT- 168
Db 122 SQYLLLVQVPPKRISSIDSVTVNBSGNTVLQMGANGREPEPTWRHLLPTGRFEGEEY 181
Qy 169 LSVSOLELDQSGTWTCTCTVLQNGKQKVEFFIDIVLAFQASSIVYKKEGEQVE--FSPEPLA 226
Db 182 LELTLGTRQSGKYEC-----KLANEVSAAVQKYQVTVNPPPT 220
Qy 227 FIVEKLTGSGEL-----WMQERASSKSKWITFDLKNKE---V 261
Db 221 ITESK---SNEAATGGRQASLKCEASGAVPADPFEWYRDDTRINSANGP--EIKSTEGQSSL 275
Qy 262 SVKRVITQDPKLOMGKKLPLHLTLPOALPOYKAGSGNLTALAEAKTGKGLHQBENVLWMPATQ 321
Db 276 TVTNVATEE-----HY--GNVYTCVAANKRGUVN--ASLVLFK- 307
Qy 322 LQKNULTCEWVGPPSPKLM---LSLKLKNEKAKVSRREKPVVNLVLEAPGMMQCILS 373
Db 308 -----PGSVRGINGISISLAV-----PMLWL---ASLILCILS 336

```

RESULT 47
S28061
SCPl protein - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #ext_change 08-Oct-1999
C/Accession: S28061
R:Meuwissen, R.L.J.; Offenberg, H.H.; Dietrich, A.J.J.; Risewijk, A.; van Iersel, M.; H
EMBO J. 11, 5091-5100, 1992
A>Title: A coiled-coil related protein specific for synapsed regions of meiotic prophase
A/Reference number: S28061; PMID:93099884; PMID:1464329
A/Accession: S28061
A/Molecule type: mRNA
A/Residues: 1-946 <MEU>
A/Cross-references: EMBL:X67805; NID:g57212; PIDD:CAA48006.1; PID:g57213
C/Genetics:
A/Gene: SCPl
C/Keywords: DNA binding

Query Match	5.2%;	Score 121;	DB 2;	Length 946;
Best Local Similarity	19.8%;	Pred. No. 3.7;		
Matches	96;	Conservative	95;	Mismatches 175;
				Indels 118;
				Gaps 24;

9 HLLVLQL----ALPATOQNKVLGKGDIVELTCTASQKSIQEH----WKSNSQIK 60

```

Db      323 HSLVLTETLAKTTCLEELLRTQGRLENNEDQLKITLWELQKSSSELEMTKFFNNKEVE      362
Qy      61 ILGNQSPFLTKGPKSLNDR-----ADRRSLMQGNPELLIKLKIED-----      103
Db      383 -LEELTKITLAEQKLLDEKQYVKLAEBLQGEDELFPLLQTRKEKIHDLDEVQVTVKTS      441
Qy      104 SDPTICEVEDEKQEEVQ--LTVFGFLTANSDPHLQOGSLT-----LTLSPPGSSPVSQC      155
Db      442 EEHYTLQVBEAKTELEKEKLNIELTANSDMLLENKLVQVBAADWMLKKGQIEDLINC      501
Qy      156 RSPGRNNIQGKTLVSQLELQDS-GTWCTVTLQNKQKVEFKDILVLAFOKASSIY--      212
Db      502 KQGEEMLQIETLEKEMNNLDELESVRKEFIQGGDEVKCLD--KSEBNARSIEYV      558
Qy      213 -KKEGG-----QVEFSPPLAFVYEKLTGSGELMWQERASSSKSWITFPLK-      257
Db      559 LKKEKQMKILLENKNNLKQKQIENK--SKNIEELHQBENKA-LKKSSSAENKQJNAAEIKV      614
Qy      258 -----NKEUSVVRVYQDPKLGQKCLPHLTLQPALPQYAGSGN      296
Db      615 NKLELELSTKQKFEEMINNYQKELEIKKISSE-KL-LGEVEKAKAVADEVKL---QKE      669
Qy      297 LTLTLEAKTGKL-----HQEVNLVVMRATQ--LQKNILCEAWGTPSPMLSTLKEN      346

```

```

Db      670 IDLCOHKIAEMVVALMEKHQYDKTVERDSEGLTYNRRPQ---OSSAKVALELEISN 726
Qy      347 -----KEAKVSKREKPVWVNLNPEAGMOCILSDSGQVLLESNIKVLP---TW----- 390
Db      727 IRNELVSLKKQLEVEKEKEKXKNQEN---TAILTDKKKQKIQASLSEPEATSMKFD 783
Qy      391 -STP 393
Db      784 KTTTP 787

```

```

RESULT 48
151669
tumor suppressor - African clawed frog
C/Species: Xenopus laevis (African clawed frog)
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C/Accession: J51669
R/RefSeq: W.E.; Reale, M.A.; Candia, A.F.; Wright, C.V.; Cho, K.R.; Fearon, E.R.
Dev. Biol. 166, 654-665, 1994
A/Title: Expression of a homologue of the deleted in colorectal cancer (DCC) gene in the
A/Reference number: J51668; MUID:95113183; PMID:7813784
A/Accession: J51669
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1427 <PIE>
A/Cross-references: EMBL:U10986; NID:g606873; PIDN:AAA70168.1; PID:g606874
C/Genetics:
A/Genes: XDCca

```

```

Query Match      5.2%; Score 121; DB 2; Length 1427;
Best Local Similarity 20.8%; Pred. No. 6.4;
Matches 94; Conservative 64; Mismatches 160; Indels 126; Gaps 24;

Qy      34 GDTVELTCTASQKSIQFHW-KNSNQIKILNQSGFLTKGPKSLNDRADRSRLMDGQNF 92
Db      154 GDTALLREITGEPMPTISWQNEEDLVK-----TPG-----DPRLLVLPST- 196
Qy      93 PLIKNLKIEDSDTYICEVD-----QKEEVQL-----VFGLTANSPTHLLOG 136
Db      197 -LQISRLQTAGGYRCLAKNPGSARVGENAEALRILSSGLHQQVF-LQRPVVAIEG 254
Qy      137 QSLTL-----TLSPGSSPSVOCSPRGKNIOGKTLVSQLELDGSGTWTCTV 186
Db      255 QDAVLECAVSGVPPPTIVMVGQDEP-VPIRT-RKSVLGGSNLLISNVTDDAAYTCVA 312
Qy      187 LQNKKEVFKIDIVLA---FQKASIIYKKEGEQVESFPLA---FTVEKLTGSGEL-- 238
Db      313 TYKNENISFSADLTVMWPPQFLNHPANLYAVESMDIEFECVSGKPSFTVMTXNGEVI 372
Qy      239 ---MWQAEKAS-----SSKSWITPDLKKEVSVKRVTO---DPKIQMGKKLPLHL 282
Db      373 PSDYQIYDGSNLRILGLVKSDEGYQCIANENGNITVQQLIIPPAVSSSILP--- 429
Qy      283 TLPPQAL-----PQYAGSGNL---TLAEAKTGKLHQEVNLVVMRATIQ- 323
Db      430 SAPRNVVLVVSREVRLSMRPPVESKGNITVTVFSGQVQREBRAVNTSQPISTLQTV 489
Qy      324 KNLICE-----VMGP-----TSPKLMSLLENKE-----AKVSKRE 355
Db      490 GNLTFEETVYNNRVVAYNMVGEGSSQEVKVVYQFLEQVPGFEVNDQVVSTPAITSVL 549
Qy      356 KPVVNLNPEAG-MWQCLSDSGQVLLESNIKV 386
Db      550 PPAVANGVGVGRLECAETFSGR---EQNIEV 578

```

```

RESULT 49
A31923
amalgam protein precursor - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change 21-Jul-2000
C/Accession: A31923

```

```

R/Seeger, M.A.; Haffley, L.; Kaufman, T.C.
Cell 55, 589-600, 1988
A/Title: Characterization of amalgam: a member of the immunoglobulin superfamily from Drc
A/Reference number: A31923; MUID:89028670; PMID:3141062
A/Accession: A31923
A/Molecule type: DNA
A/Residues: 1-333 <SEE>
A/Cross-references: GB:M23561; NID:g156920; PIDN:AAA28367.1; PID:g156921
C/Genetics:
A/Genes: FLYBase:Ama
A/Cross-references: FLYBase:Fbgn0000071

```

```

Query Match      5.2%; Score 120.5; DB 2; Length 333;
Best Local Similarity 19.4%; Pred. No. 0.97;
Matches 56; Conservative 48; Mismatches 123; Indels 61; Gaps 9;

Qy      10 LLVLQAL-----LPAATGKNVVLGKDDYELTCTASQKSIQFHW-----KNSNQ 58
Db      10 LIFCLALSDLSVLSAPVLSQISKDVVASVGSVEFNCTVEVQQLSVSMKRPSESDTNS 69
Qy      59 I-----KILG-----NQSGFLTKGPKSLNDRADRSRLMDGQNFPLIKNLKIEDSDTYIC 109
Db      70 VLISMNRLSLPKRKNVTVTEGPKT-----GSAITFRQNIENVDMGPEYC 117
Qy      110 EV-----EDQKEVQL---VFGLTANSPTHLLOGSITLTLESPPGSSPSV----- 153
Db      118 QVAVSATEKVTKKLSIQIKTPPIAENTPKSTLVTEGQNLCTHANGFPKPTISWAREH 177
Qy      154 QCRSPRGKNIOGKTLVSQLELDGSGTWTCTVQLQNKKEVFKIDIVLA-FQKASIVYK 213
Db      178 NAWMPAGGHILAEPTLIRSHRMDRGGYCIANQNGQDPDKLIRVEVFRQIAVQRP 237
Qy      214 KEQGEVFSFPLAFTVEKLTGSGELWQ-----AERASS 248
Db      238 KIQMWSHSALECSVQGYAPFVWVHKNGVPLQSSRHHEVNAVTAASS 285

```

```

RESULT 50
ORRG
secretory component precursor - rabbit
N/Alternate names: poly-Ig receptor; polymeric immunoglobulin receptor
N/Contains: free secretory component; transmembrane secretory component
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 15-Nov-1984 #sequence_revision 15-Nov-1984 #text_change 16-Jul-1999
C/Accession: A02111; A28077
R/Moskov, K.E.; Friedlander, M.; Blobel, G.
Nature 308, 37-43, 1984
A/Title: The receptor for trans epithelial transport of IgA and IgM contains multiple imm
A/Reference number: A02111; MUID:84142246; PMID:6322002
A/Accession: A02111
A/Molecule type: mRNA
A/Residues: 1-773 <MOS>
A/Cross-references: GB:X00412; GB:K01291; NID:g1595; PIDN:CAA25118.1; PID:g1596
A/Note: the authors translated the codon ACC for residue 54 as Asn
R/Fructiger, S.; Hughes, G.O.; Hanly, W.C.; Jaton, J.C.
J. Biol. Chem. 263, 8120-8125, 1988
A/Title: Rabbit secretory components of different allotypes vary in their carbohydrate co
A/Reference number: A28077; MUID:88228032; PMID:3131339
A/Accession: A28077
A/Molecule type: protein
A/Residues: 87-114;410-424 <FRU>
C/Comment: This receptor binds polymeric IgA and IgM at the basolateral surface of epithe
process, cleavage occurs to separate the extracellular portion, also known as the secret
C/Comment: The five domains exhibit homology with immunoglobulin V regions. The similarit
C/Comment: Alternative splicing in the extracellular domain leads to high or low molecule
C/Superfamily: secretory component; immunoglobulin homology
C/Keywords: alternative splicing; duplication; glycoprotein; immunoglobulin receptor; pol
F/1-18/Domain: signal sequence #status predicted <SIG>
F/19-773/Product: transmembrane secretory component #status predicted <MATW>
F/19-575/Product: free secretory component #status predicted <EXT>
F/30-647/Domain: extracellular #status predicted <EXT>
F/39-117/Domain: immunoglobulin homology <IM1>
F/148-227/Domain: immunoglobulin homology <IM2>

```

F:253-326/Domain: immunoglobulin homology <IM3>
 F:362-440/Domain: immunoglobulin homology <IM4>
 F:471-540/Domain: immunoglobulin homology <IM5>
 F:648-670/Domain: transmembrane #status predicted <TM>
 F:671-777/Domain: intracellular #status predicted <INT>
 F:46-115,155-223,260-324,369-438,478-538/Disulfide bonds: #status predicted (partial) #status experimental
 F:108/Binding site: carbohydrate (Aen) (covalent) #status experimental
 F:418/Binding site: carbohydrate (Aen) (covalent) #status experimental

Query Match 5.2%; Score 120.5; DB 1; Length 773;

Best Local Similarity 20.3%; Pred. No. 3;

Matches 108; Conservative 75; Mismatches 199; Indels 149; Gaps 26;

```

QY 5 VPFRIILL-----VLQALALPAATQGN-----KVLGKKGGPTVELTCAQSKKSLQ 50
DB 210 VTIKHLQNDAGQVYQSGSDPTAEQNVDLRLTPGLLYGNLGGSVTFECALDSEDAVA 269
QY 51 FHMKNSTQIK-----ILNGGSFPLTKGPSKLANDRADSRSLW---DQGNFPLIYKLIKIED 103
DB 270 V--ASLRQVRGGNVVIDSGTI-----DPAFEGRLFPKAEHGHSVVIAGLRKED 318
QY 104 SPTTYICEVEDQKEEVQLLVFGLTANSDFHLLQ---GQSILTLLEBP-----FGSSPSVQC 155
DB 319 TGNVLCGVQSN-----GQSGDGPQLRLQLFVNEBIDVRSPPVLKGFPGSGSVTTRC 369
QY 156 -RSPR-----GKNI-----GGKTLTV--SOLE 175
DB 370 PYNPKRSBSHLQLYMEGSGTTHLLVDSGEGLVQKDYGRLLAFEPENGTFSVLANQTT 429
QY 176 LQDSGTWTCTVLQONKQVEFKIDIVLAFQKASSI--VYKKEGQVRS--FPLAFTEK 231
DB 430 AEDEGFYWC-VGDDDESILTSVKLQIVDGEPSPTIDKFAVGEVEITCHPCKY---- 484
QY 222 LTSGSELMQAEARASSSKSWITF-DLKKEVSVKRVTDPKLQMGKULPHLITLPOALPQ 290
DB 485 -----FSEKTYCKMNHGCELDPTKLISSSGDLVYCKNNNVLTLLTDSVED 531
QY 291 YAG---SGNLTALAEKTKGLHQEVNLVYMRATOLQKNLTCEVWGPPTSFKMLSLKLNK 347
DB 532 DEGWYWC-----AKDGEFEVAVAVVELTEPAKVAVEPAKYPVDDAKAAPPAEAK 584
QY 348 EAKVSKREKPVVNLPEAGMOCCLSDSGQVLLSENIKVLPTWSTPVHPRASALPAPPTG 407
DB 585 ----AKACFPV----PRRQWYPL-----SRKLRR-SCP-ERLLAEVAVQGS 622
QY 408 SALPPOPTASALPDPAPASALPALAVISFLIGLGI-----GVACVLARTR 453
DB 623 AEDPASGSRASVDASSAGSGSAKVLISTVLPLGLVLAAGAMAVAIARAR 673

```

RESULT 51

T18358
 apolipophorin precursor protein - tobacco hornworm
 C:Species: Manduca sexta (tobacco hornworm)
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
 C:Accession: T18358
 R:Sundermeyer, K.; Hendricks, J.K.; Praead, S.V.; Wells, M.A.
 Insect Biochem. Mol. Biol. 26, 735-738, 1996
 A:Title: The precursor protein of the structural apolipoproteins of lipophorin: cDNA and
 A:Reference number: Z18891; MUID:97166603; PMID:9014323
 A:Accession: T18358
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3305 <SUN>
 A:Cross-References: EMBL:U57651; NID:g1399217; PID:g1399218; PIDN:AAB53254.1

Query Match 5.2%; Score 120.5; DB 2; Length 3305;

Best Local Similarity 22.5%; Pred. No. 21;

Matches 97; Conservative 66; Mismatches 176; Indels 93; Gaps 21;

```

QY 26 NKVVLGKKGDVLTCTASQKKS-----IQFHMKNSTQIKILNGGSFPLTKGPSKLANDRA 80
DB 1852 NSVVVDADGRVYKIDSSIVLSKAHPVLDIQVHSPSSDKIRRLYLQSSLSSTQGLTEKVV 1911

```

```

QY 81 DSRSLMDQGNPLIKLIKIEDSDTYICEVEDQKEEVQLLVFGLTANSDFHLLQGSILT 140
DB 1912 DN-----INDICLD---AVSEANVQKDNV---AFKVVANAKELGMKNNGID 1951
QY 141 LTLESPPSSPSVQCRSPR-SKNIQSGKTLTSLVSOLELDS---GTWCTVLQONKQVEFK 196
DB 1952 IS-SKDSGGRLEPHANNDKNVLSGSTSPISNQGOKTIIIESSGSYKVEEQKSNFX 2010
QY 197 IDIVLAFQKASSIYKKEGEO-VFSPPLAFVTEKLTGSELMQAEARASSSKSWIT-F 254
DB 2011 Y-----IRVFTDTSNKEGVETFPFNVAL-----GERSYVAESRVYNY 2046
QY 255 DLKKEVSVKRVTDPKLQMGKULPHLITLQALPQVAGSNLTALAEKTKGLHQEVNL 314
DB 2047 EYKNSYVCEERKQCAHAIEQSK--IDWSTGMVNNVINA-LDL--RKLGV----- 2093
QY 315 VYMRATOLQKNLTCEVWGPPTSFKMLSLKLNKAEKAVSKREKPVVNL-PEAGMOCCLLS 373
DB 2094 ----APELGLQMRDEVSDRRPRFTLDLHT-NKE-----DRKYHLAAVYTPENGHY----A 2140
QY 374 DSGQVLLSENIKVLPTWSTPVHPRASALPAPPTGSAALPD-----POTASALPDPAA 425
DB 2141 SGVTVRLPSRYMALE--YTLTHPTSODLPFPKGEACIDLDKNRPGHKTSAFLVDYSNS 2198
QY 426 SALPALAVISF 437
DB 2199 GSEDKAVAEIGF 2210

```

RESULT 52

hypothetical protein DKFZps6411922.1 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 23-Jul-1999
 C:Accession: T08678
 R:Mambuti, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, May 1999
 A:Reference number: Z16469
 A:Accession: T08678
 A:Molecule type: mRNA
 A:Residues: 1-584 <MAN>
 A:Cross-References: EMBL:AL049946
 A:Experimental source: fetal brain; clone DKFZps6411922
 C:Genetics:
 A>Note: DKFZps6411922.1

Query Match 5.1%; Score 119; DB 2; Length 584;

Best Local Similarity 20.3%; Pred. No. 2.6;

Matches 69; Conservative 57; Mismatches 130; Indels 84; Gaps 17;

```

QY 12 LVTLQALALPAATQG-----NKVVLGKKGDVLTCTASQKKSIOFHMKNSTQIKILGNQ 65
DB 285 LILQVLLEPMEKDFHDPISSEKITAMAGHTISLNCAGTPPSLVW-----VLPN- 336
QY 66 GSFLTKRP--SKANDRADSRSLMDQGNFPLIYKLIKIEDSDTYICEVEDQKEEVQLLV- 122
DB 337 GTDLQSGQGLQRFYHKADGM-----LHISGLSSVDAAGAVCARNAAGHTERLVS 386
QY 123 --FGL--TANSDTG-----LLQGSILTLLESPSSPSVQCRSPRGNKIQGK----- 167
DB 387 LKVLKEPASQYNNLVSIINGETLXPCPTPRGAGQGRFSTYLPNGNHLBEPQLGKAVSL 446
QY 168 ---TLASVQLELDQSGTWCTVLQONKQVEFKIDIVLAF-----QKASSIYKKEGEO 218
DB 447 LDNGTLTVREASVDRGTIVYCRMETEYGPSVTSIPVIVIAVPPRITSEPTPIVITRGNT 506
QY 219 VEFSPPLAFVTEKLTGSELMQAEARASSSKSWITPPLDKKEVSVKRVTDPKLQMGKUL 278
DB 507 VKLNC-MAMGIPK---ADITWELPDKSHLAAVQOARL-----YGNRF 544
QY 279 PLH-----LTLPOALPQVAG-----SGNLTALAEKTKGLH 309

```


Db 545 -LHPGSLTIQATCRDAGFYCKMAKNI-LGSDSKTYIYH 582

RESULT 53

JC4917

Signal transducing adaptor - mouse
C:Species: Mus musculus (house mouse)

C>Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 21-Jul-2000

C:Accession: JC4917

R:Takeishi, T.; Ariga, T.; Asao, H.; Tanaka, N.; Higuchi, M.; Kuroda, H.; Kaneko, K.; M

Biochem. Biophys. Res. Commun. 225, 1035-1039, 1996

A:Title: Cloning of a novel signal-transducing adaptor molecule containing an SH3 domain

A:Reference number: JC4916; MUID:96374438; PMID:8780729

A:Accession: JC4917

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-548 <TAK>

A:Cross-references: GB:U43900; NID:G1556460; PIDN:AA052840.1; PID:G3645912

A:Experimental source: T cell

C:Comment: This protein contains a Src-homology 3 domain and the immunoreceptor tyrosine

C:Genetics:

A:Map position: 2A2-B

C:Superfamily: SH3 homology

F:2-540/Product: signal transducing adaptor molecule #status predicted <MAT>

F:217-264/Domain: SH3 homology <SH3>

F:359-387/Region: immunoreceptor tyrosine-based activation motif

Query Match 5.1%; Score 118; DB 2; Length 548;

Best Local Similarity 21.0%; Pred. No. 2.8;

Matches 92; Conservative 62; Mismatches 174; Indels 110; Gaps 19;

Qy 69 LTKGPKNDKADSRSLM--DOGNP-----LTKKLTIEDSDTYIC--EVEDQKEE 117

Db 97 LNKHPKCEKCKLALMVEWTEDEPKNDPQLSLISAMIKLK--EGQVTPFAISQAAEQAKA 155

Qy 118 VQLVLF--GLTNSDTHLQGGSLTLTLSPGSSPSVQCRSPRGKNI-----QGQKT 168

Db 156 SPALVAKDPGVATVATKEEDLAKAIELSKQRQOSAPVSTLYPSTSLTLTNHQHGKRV 215

Qy 169 LSVGSLLELDGSGTCTVLQOKKVEFKIDIVLAFOQASSIVYKKEGEQVEFSPLAFT 228

Db 216 RAVVDFAEDNELT-----FRAGELITVLDDSDPNMKGETHQVGLFSPNFV 264

Qy 229 VEKLTGSGELMWQAEKASSKSMITFDLNKEVSVKRTQDPK-----LQWKG 276

Db 265 TADULTA-----EPERITEKKTQVF--NDVQIETIEPEPEPAFTEDKMDQLQW-- 313

Qy 277 KLPLHLT-----LQALPQVAGSGNLTLAEKTKGLHQ-----EVNLVYMRATQL 322

Db 314 --LQSTDPDNDQPLPELHLAEWCQMGPLIDKEDIDRKISELSELNVKWEALSL 370

Qy 323 QKNLTCEWGTSPKLTMLSTLKENKEAKVSKREKVVWLANPEAGMWQGLSDSQVLLS 382

Db 371 YTKLWNE--DPMYSWYAKLQSQVYLLQSSAVASGVYRGP-----AQSGTYLVAG 418

Qy 383 NIKV--LPTWSTPVPRAS-----ALPAPPTGSAALPDPTASALPD----- 421

Db 419 SAQWHLDSYSLPPEQLSSISOGAVPS--SANDALPQOOTQASYPANAVSVQGSYPSQA 477

Qy 422 ---PPAASALPALAVT 435

Db 478 SIYSPPAAMAAAAAAAV 495

RESULT 54
T29757
Protein UNC-89 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C:Accession: T29757
R:Du, Z.; Le, T.T.; Wilson, R.
Submitted to the EMBL Data Library, May 1997

A:Description: The sequence of C. elegans cosmid C09D1.

A:Reference number: Z20679

A:Accession: T29757

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-6642 <DU2>

A:Cross-references: EMBL:AF003131; PIDN:AA054132.1; GSPDB:GNO00019; CESP:unc-89

A:Experimental source: strain Bristol N2; clone C09D1

C:Genetics:

A:Gene: CESP:unc-89

A:Map position: 1

A:Initons: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 63

/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match 5.1%; Score 118; DB 2; Length 6642;

Best Local Similarity 18.4%; Pred. No. 80;

Matches 94; Conservative 67; Mismatches 145; Indels 206; Gaps 22;

Qy 32 KKGDVLELTCTASOKSIQFHW-KNSNQIKILGNQGSFLTGPKSKLNDKADSRSLWDQG 90

Db 2899 KKGSAVFECVVDPTGVCCKMLKDGEIELIARIVQRTGP-----EG 2943

Qy 91 NF--PLTKKLTIEDSDTYICEV-----DQKE-----VQLLVFG 124

Db 2944 HITQELVDVTPBEDAGKTCIYENTAGKDCATLTVISLEKSEKKAPEFVLQDK 3003

Qy 125 LTANSDDTHLQGGSL-----TLFLSPGSSPSVQCRSPRGKNIQGGKTLVS 172

Db 3004 TTTSKENVLECKVIGEPKPKVSMHDNVSREKPNSEKTIITQESTIVESVEGVERVIT 3063

Qy 173 QLELDGSGTCTVLQOKKVEFKIDIVLAFOQASSIVYKKEGEQVEFSPLAFTVEKL 232

Db 3064 SSELSHQKVTG-----IAENTGTSKTEAF-----L 3090

Qy 233 TGSSELMLWQAEKASSKSMITFDLNKEVSVKRTQDPKQKMKLPLHLT--PQALP 289

Db 3091 TVQGE-----APVETKELQNKELSI-----GBKVLSCSVGSPQPHV 3128

Qy 290 QVAGSGNLT-----LAEAKTGKLEQVN--LVWMRATQ----- 321

Db 3129 DFFSFTTIVETIKTISSSIALE-----HDQTNHMRVVISQITIEDIVSKATATNS 3182

Qy 322 ---LQKNLTCEWGP-----TSPKLTMLSKLENK-----EAKVSKREKV--- 358

Db 3183 IGTATSTSKITTKVPAVFPQGLKJTSVKEKEIKNEVKVGSAPVDFPKDPVSEDG 3242

Qy 359 ---WVNLPEAGM-----QCLSDSQVLLSNIKVLPTSTVHPASALPAPPTGSA 410

Db 3243 NHEKKKNPEGVFTLVVKQATTDAGKYTAKA-----SNPAGTAAESSAAEAETQS-L 3293

Qy 411 PDPQ-----TASALPDP 422

Db 3294 EKPFVARELVTEVKINETAITLSVTVKGVDP 3325

RESULT 55
IURTNC
neural cell adhesion molecule short domain form precursor - rat
N:Alternate names: NCAM-140
C:Species: Rattus norvegicus (Norway rat)
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 22-Jun-1999
C:Accession: S00846; B37795; I58136
R:Small, S.J.; Shull, G.E.; Santoni, M.J.; Akeson, R.
J. Cell Biol. 105, 2335-2345, 1987
A:Title: Identification of a cDNA clone that contains the complete coding sequence for a
A:Reference number: S00846; MUID:88059265; PMID:3680385
A:Accession: S00846
A:Molecule type: mRNA
A:Residues: 1-858 <SWA>
A:Cross-references: EMBL:X06564
R:Small, S.J.; Akeson, R.
J. Cell Biol. 111, 2089-2096, 1990
A:Title: Expression of the unique NCAM VASE exon is independently regulated in distinct

A:Reference number: A37795; MUID:91035620; PMID:1699951
 A:Accession: B37795
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 340-381 <SM2>
 R:Small, S.J.; Haines, S.L.; Akesson, R.A.
 Neuron 1, 1007-1017, 1988
 A:Title: Polypeptide variation in an N-CAM extracellular immunoglobulin-like fold is dev
 A:Reference number: 158136; MUID:90166485; PMID:2483093
 A:Accession: 158136
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 355-364 <RES>
 A:Cross-references: GB:M32611; NID:9205643; PIDN:AAA1679.1; PID:9205644
 C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
 C:Comment: Various forms of NCAM are produced by alternative splicing.
 C:Gene: NCAM
 C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
 C:Keyword: alternative splicing; brain; cell adhesion; duplication; heparin binding; st
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-858/Product: neural cell adhesion molecule, short domain form #status predicted <M
 F:20-721/Domain: extracellular #status predicted <EXT>
 F:34-98/Domain: immunoglobulin homology <IMM1>
 F:132-191/Domain: immunoglobulin homology <IMM2>
 F:152-156/Region: heparin binding #status predicted
 F:161-165/Region: heparin binding #status predicted
 F:228-290/Domain: immunoglobulin homology <IMM3>
 F:263-272/Region: NCAM binding #status predicted
 F:323-398/Domain: immunoglobulin homology <IMM4>
 F:430-492/Domain: immunoglobulin homology <IMM5>
 F:529-606/Domain: fibronectin type III repeat homology <FN3A>
 F:635-695/Domain: fibronectin type III repeat homology <FN3B>
 F:740-858/Domain: intracellular #status predicted <INT>
 F:41-96/Domain: intracellular #status predicted <INT>
 F:41-96,139-189,235-288,330-396,437-490/Disulfide bonds: #status predicted
 F:222,316,348,434,460,489/Binding site: carbohydrate (Aaen) (covalent) #status predicted

Query Match 5.1%; Score 117.5; DB 1; Length 858;
 Best Local Similarity 19.1%; Pred. No. 5.5;
 Matches 79; Conservative 52; Mismatches 143; Indels 139; Gaps 17;

3 RG-VFRLHVLVLAALPAATGKNNKVVLGKGDVLELTCTASQKKSIOFHH-KNSNDIX 60
 196 KGEINFKDIOVIVNVPPTVQAQOSIVNATANIAGSVTLVCDAGPPEPTMSWTKDGEIE 255
 61 ILNGSSFLTKGPRSLNDRADSRSLMDQGNFLLIKLIKIEDSDTYICEV---DQKE 116
 256 -----NEEEDDEKHIIFSDSSELTITRVNDKNDABEYCIANKAGEDDA 299
 117 EVQLLVFG---LTANSDTHLQGSLLTTLSPGSSPSVOCRRPRGKNIOGKTLVS 172
 300 SIHLVFAFKITTYVENQTMELERQVTLTCEASGDPPISTWRT-----STR 347
 173 QLELDSGTWTCTVLOKQKVEFKIDIVLAFOKASIIYKKEGROVESPPFLATVEKL 232
 348 NISSEKKSMT---RPEKQETLDGHMVRSHARVSSLTK-----SIOT----- 388
 233 TGSSELIMQAEKASSSSKSWITFDLNKKEVSVKRVVODPRLQMGKPLHLTLPLQALPOYA 292
 389 TDAGSYICTASNTIGQDS-----QSMYLEVGYPRLQO-----PVAV---YT 427
 233 GSGNLTALAEATGKLGHOENVLVWRATQLOKLTCEVWGSPKLMISLKLKNEAKYS 352
 438 WEGN-----QVNITCEVFAVPS-----ATIS 448
 353 KKEKPVVVLNPEAGMWQCLSDSGOVLESNIKVLPTSTVHPRASALPAP 405
 449 -----WFRD-----GQLPSSNYSNIKIYNT---PANSYLEVTP 479

RESULT 56
 T30581

neural cell adhesion molecule L1.1 - zebra fish (fragment)
 C:Species: Brachydanio rerio (zebra fish)
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
 C:Accession: T30581
 R:Tongioigi, E.; Bernhard, R.R.; Schachner, M.
 J. Neurosci. Res. 42, 547-561, 1995
 A:Title: Zebrafish neurons express two L1-related molecules during early axonogenesis.
 A:Reference number: 220875; MUID:96155762; PMID:8566941
 A:Accession: T30581
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1197 <TON>
 A:Cross-references: EMBL:X89204; NID:91065713; PID:91065714; PIDN:CAA61490.1
 C:Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homology; i

Query Match 5.1%; Score 117.5; DB 2; Length 1197;
 Best Local Similarity 18.0%; Pred. No. 8.6;
 Matches 83; Conservative 84; Mismatches 122; Indels 123; Gaps 19;

9 HLLVLVLAALPAATGKNNKVVLGKGDVLELTCTASQKKSIOFHHKNSNOIKIINGQSF 68
 259 HYTVVEAA--PYWTRSPFEHLVAPGETVRLDCAADGIPAPNITW--SINGVPVSGTD--- 312
 69 LTKGPRSLNDRADSRSLMDQGNFLLIKLIKIEDSDTYICEVDQKEEVQLVFGLTAN 128
 313 -----VDPRRRV---SSGKLISNVFSDTAVYQCEAVNKGISILI-----N 351
 129 SDTHLQGSLLT---LESPPGSSPSVOCRS-----PRGKNIOGK 167
 352 THVHVEVLAQILTPDERLQATAGQVLMDCRTFGSPLPKIHWEILDSIPALNNAKISQ 411
 168 T-----LSVQLELDSGTWTCTVLOKQKVEFKIDIV-----VLAFOKASIIY 212
 412 TTNGSLKISNVSEDSNRKTVSVSTNKSISADVEVLAKRTYIGVPPONLHVIKRSDALH 471
 213 KKEGEVFEFPLAFVTEKLTGSGELIMQAEKASSSSKSWITFDLNKKEVSVKRYTODPKL 272
 472 CK-----YTVDHNLKSPVQKNGKHITNS--TSNKKYHIEBSLAKLVLDVQM 517
 273 Q-MGKPLRLHLTLPLQALPOYAGSGNLTALAEATGKL-----HOENVLVWRATQLO 323
 518 EDWG-----IYSCVSTLDSPTASGYITVDKRPPOSILKI-----SEKGE 559
 324 KNLCEVWGPT---SPKLMISLKLKNEAKYSREKRVWVLNPEAGMWQCLSDSGQVL 379
 560 KSVTIS--MPSVENNSPTEYVLENBGE-----TDBEGMQKRSVSQDID 605
 380 LESNIKVLPTSTVHPRASALPAPPTGSALPDPQTASALPD 421
 606 SMRSICYSKYNHQLRAVNSIGTAPTESSL-SYSTRAKRD 646

RESULT 57
 T43027
 neural cell adhesion molecule L1 - goldfish.
 N:Alternate names: E587 antigen
 C:Species: Carassius auratus (goldfish)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
 C:Accession: T43027
 R:Giordano, S.; Laessle, U.; Lottepeich, F.; Stuenkel, C.A.O.
 submitted to the EMBL data library, April 1996
 A:Description: Molecular cloning of goldfish E587 antigen, a cell adhesion molecule expr
 A:Reference number: 222294
 A:Accession: T43027
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1232 <GIO>
 A:Cross-references: EMBL:U55211; NID:91305526; PID:91305527; PIDN:AAA9159.1
 C:Superfamily: neural cell adhesion molecule L1; fibronectin type III repeat homology; i
 C:Keywords: cell adhesion; membrane protein

Query Match 5.1%; Score 117.5; DB 2; Length 1232;
 Best Local Similarity 21.5%; Pred. No. 8.9;

Matches 102; Conservative 68; Mismatches 173; Indels 131; Gaps 25;

QY 2 NRGVFRLHLVLQALPAATQGNKVVLGKKGDVLTCTASQKKSIOFHMKNSNOIKI 61
 Db 286 SGGSVGHVYATVEAA--PWTTRPRENLVAPGEVTRLDQCAEGIPFTNITW--SMNGAPI 342
 QY 62 LGNGGSFLTKGPKSKLNDADSRSLMDQGNFPLIKLKIEDSDTYICEVDQKEEVL 121
 Db 343 AGTD-----PDRRHV---SSGTLILTDVQI--SDPAVYHVEATNKHGNI 383
 QY 122 VEGLTANSDTHL--LQGSGLT--LTLESPGSSPVQCR--SPRGK----- 161
 Db 384 I-----NTHVAVLEPRLITFEDDLKYATGQVTLDCRRFGSGFQKVMQITNSGAL 438
 QY 162 -NIQSGK-----LSVSOLEDSGTWTCTVLQONKQVEFKIDIVLAFQKASSIVYKKG 216
 Db 439 ANAKMSQTSDDGLQISDVSEBSSMYTCSV--STSMISALVYL---NNTKIVDPQD 493
 QY 217 EQV---EFSPPLATVTEKLTGSGELMWQAEBAASSKSWITFDLKNKEVSKRVTDPKL 272
 Db 494 LRVLRGDAVLQCRVTVHMLKQPTIQMKDKHKITSS-----ANDDKY 537
 QY 273 QMGKKLPLHLTLPAQLPVAGSGNLTLLAEKTKGLHQEVLVYWRATQLOKLTCEWGS 332
 Db 538 TESPFGSKIT---DVQMEDSGIYCEIISTLSDVSATGSIW-----LKK----- 580
 QY 333 PTSP-KLMLSLKEN-----KEAKVSRERKRPVWVWLNPEAGMOCCLSDS 375
 Db 581 PGSPHSLSLSEKKEKSVTLSTMWPGAENNSPISEVIERKENQ---NGKQHMEEYRVP 636
 QY 376 GQVL-LESNIXVLPWSTFVHPRASAL-----PAPPTGALPDPQTASALPD 421
 Db 637 QDITLHLEIHLQPSYTY---HFRVKGVNGIGMSSESPSESY---STPAKPD 682

RESULT 58
 A32579
 neuroglian - fruit fly (Drosophila melanogaster)
 CSpecies: Drosophila melanogaster
 CDate: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 CAccession: A32579
 R|Bieber, A.J.; Snow, P.M.; Horroch, M.; Patel, N.H.; Jacobs, J.R.; Traquina, Z.R.; Schib
 Cell 59, 447-460, 1989
 ATitle: Drosophila neuroglian: a member of the immunoglobulin superfamily with extensively
 AReference number: A32579; MUID:90030418; PMID:2805067
 AAccession: A32579
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1239 <BIE>
 A:Cross-references: GB:M8231; NID:G157998; PID:AAA28728.1; PID:G157999
 A>Note: the authors translated the codon TAT for residue 1234 as Thr and AAA for residue
 C:Genetics:
 A:Gene: FlyBase:FlyBase:FlyBase:FBgn0002966
 A:Cross-references: FlyBase:FBgn0002966
 C:Superfamily: neural cell adhesion molecule I; fibronectin type III repeat homology; i
 C:Keywords: alternative splicing; cell adhesion; duplication; membrane protein
 F:353-412/Domain: immunoglobulin homology <IMM2>
 F:446-502/Domain: immunoglobulin homology <IMM3>
 F:535-596/Domain: immunoglobulin homology <IMM3>

Query Match 5.0%; Score 116.5; DB 1; Length 1239;
 Best Local Similarity 19.5%; Pred. No. 10;
 Matches 65; Conservative 59; Mismatches 131; Indels 79; Gaps 14;

QY 25 GNKVVLGKKGDVLTCTASQKKSIOFHMKNSNOIKI LGNGGSFLTKGPKSKLNDADSR 84
 Db 233 GNKVVLDQNGKHPRVQVRSRQSLALRGKMELEFCYGG-----TPLPQVWWSK-DGQR 286
 QY 85 SLWD-----QGNF---PLITKLIKIEDSDTYICEVDQKEEVLVFGLTANSDTHLQGS 138
 Db 287 IQMSRRIQGHYKSLVLRQTNFDDAGTYTCDVNSGVNGAQSFSIILVNSVPPFTKEPE 346
 QY 139 LTLTLESPGSSPVQCRS-----PRGKNIQGK-----TLVSOLELQ 177

```

Db          347 IATAED---EEVVEECRAAGVPPEPKISWINGKPIEGSTPNERRRTVTQNTIRIINLVKG 403
Oy         178 DSGWTCTVLQN----OKKVEFKIDIVLAFQAKASSIVYKKGEQVFFSPLAFTVE--KL 232
Db          404 DTGNVCNAATSLGVIYVDVLIYNQAEPPTISEAPAASVTSDGRNV-----TICRV 455
Oy         233 TGSGLMWAQBARRASSKSMT--FDLK-NKEVSVKRVTQDPKLQMGKKLRLHLTLFQAL 288
Db          456 NGSPPFLVKMLRAS---NWLTCGRYVNQANGDLEIDVT----- 491
Oy         289 FOYAGSGNLTLAEKTGKLNHOEVLVVMRAIQL 322
Db          492 --FSDAGRKYTCYAQNKFGEIQADGSLVVKHEHTRI 523

RESULT 59
S49010      embryonic receptor kinase - mouse
C|Species: Mus musculus (house mouse)
C|Accession: S49010
C|Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 24-Sep-1999
R|Choi, K.; Wall, C.; Hamraty, R.; Keller, G.
OncoGene 9, 1261-1266, 1994
A|Title: Isolation of a gene encoding a novel receptor tyrosine kinase from differentiat...
A|Reference number: S49010; MUID:94181281; PMID:813130
A|Accession: S49010
A|Status: preliminary
A|Molecule type: mRNA
A|Residues: 1-1330 <CHO>
A|Cross-references: EMBL:X78568; NID:G510664; PIDN:CA55311.1; PID:G510665
C|Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homoloG
C|Keywords: ATP
F|823-1158/Domain: protein kinase homology <KIN>
F|831-839/Region: protein kinase ATP-binding motif

Query Match           5.0%; Score 116.5; DB 2; Length 1330;
                        Beet local similarity 19.7%; Pred. No. 12;
Matches 91; Conservative 72; Mismatches 167; Indels 133; Gaps 21;

Oy         33 KGDIYELICTASQKSIQFMKNQSNQIKLGNGSGTLTGPSLTNDRADRRRSIMDGNF 92
Db          568 EGEDIKLSGVN---KFLYRDITWILRLTYNNRTMHNSISKNGHSSYSI---TL 617
Oy         93 PLIIKNLIKIEDSDTYICEV-----EDQKEVOLLVFGLTANSDTHLQGSLTLTLESF 146
Db          618 NLIVIKNVLEBSGTACARNIYTGEDIRKTEVLVRDESA---PHLLQULS---DYEVS 671
Oy        147 PGSSPVOCRSPRG---KN---IQ-----GGKTLVSYOLELDGSTWCTT 185
Db          672 ISGSTTLDCQA-RGVAPQITWLKNNHKIQGPFGIILFGNSTLTFIERVTEDEGVRCR 730
Oy        186 VLQNKRYE-----FKIDIVLAQKASSIYY-----KKEGEYEF 221
Db          731 STNQGAVERAAUYLTVGOSTSDKSNLELITLYTCAVATFWMLLTFLIRKLRSSSEVKT 790
Oy        222 SF-----PLAFTYEKLTSGBELW-WQAERASSSKSM-----ITFDLKN- 258
Db          791 DYLSIMPDEVPRLEDQCEBRLPYDASNWFABERLKGSLSGAGAGKVQAFAFGIKS 850
Oy        259 ---KEVSVKRYTDQDKLQMGKKLRLHLTLFQALFOYAGSGNLTLAEKTGKLNHOEVLV 315
Db          851 PTCRVVAVKMKLEGATASEYKALMTELKILTIHGHNLVNVLNLAGACTKQGGPL-----MV 905
Oy        316 VMKATQLOKNLTCLEVWGSPSPKLMSLKLENKAKEYKREKRVYVNLNPEAGMQCLSDS 375
Db          906 IVE-----YCK-YGNLSNYL-----KSKRIDFCINKDAALHMEIKES 942
Oy        376 GOVLESNIKVLPTWSTPVHPRALPAAP-TGSALLPDQOTAS 417
Db          943 LEPGLEGGCK-----PRLDSVVSSSVTSSFPEDRSVS 975

```

```
RESULT 60
JC1509
biliary glycoprotein E - mouse
C:Species: Mus musculus (house mouse)
C:Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 07-May-1999
C:Accession: JCI1509
R:McCuall, K.; Rosenberg, M.; Nedellec, P.; Turbide, C.; Beauchemin, N.
Gene 127, 173-183, 1993
A:Title: Expression of the Bgp gene and characterization of mouse colon biliary glycopro
A:Reference number: JCI1505; MUID:93273228; PMID:8500759
A:Accession: JCI1509
A:Molecule type: mRNA
A:Residues: 1-458 <MCC>
A:Cross-references: GB:X67280
C:Comment: This protein is expressed at the cell surface and plays a determinant role in
C:Genetics:
A:Gene: Bgpe
C:Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi
C:Keywords: glycoprotein; receptor
F:1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F:160-219/Domain: immunoglobulin homology <IMM1>
F:254-303/Domain: immunoglobulin homology <IMM2>
F:339-396/Domain: immunoglobulin homology <IMM3>
F:87,104,148,199,206,210,226,258,290,294,304,333,375/Binding site: carbohydrate (asn) (c
Query Match 5.0%; Score 116; DB 2; Length 458;
Best Local Similarity 19.6%; Pred. No. 3;
Matches 100; Conservative 76; Mismatches 185; Indels 148; Gaps 23;
QY 1 MNRG-VPRHLLLVLOL-----ALPPATQGNKVLGKKGDTVELTCTAS 44
DB 8 LHKGVPMVGGLLTSLASLWSPPTTAETIEAVPQVAEDNNVLLVHNLPALGA--- 64
QY 45 QKKSIOGFHWKNSNQI-----KILGNQGSFLTGPSPKLANDRADRSLSMDQGNPLIIT 96
DB 65 -----FAMYKGNPVTNAEIVHVGVTGNTKT--TTGPAH-----SGRETVYVSNGS--LLI 109
QY 97 KNLKIEDSDTYICEVEDQ-----KEEVQ-----LLVFGLTANSPTHLQOGSLTTLTESP 145
DB 110 QKVTYKDTGVYTIETMDENFRRTATVQFHNHPLLKPNITSNNSNPVEGDDSVSLTCDSS 169
QY 146 PPGSSPVQCRSPRGKNI-----OGSKTLSVSOLELQDSGTWCTCTVLQNKQKVE-- 194
DB 170 YTDPNITVYLSMRNGBESLSEGRDLKLSBGNRTLLTNAVTRNDTGPVCEI--RNPVSVNRS 228
QY 195 --FKIDIV-----VLAFOKASSIYKKKEGQEVESFPPLAFTVEKLTSGGELMW---QAE 243
DB 229 DPFSLNIIYGPDPPII-----SPSDIYLHPGSNLTNSCHAA-----SNPPAQYFWLINEXR 279
QY 244 RASSKSMI-----TFDLKNKEVSVKRVTPQDPKLGKGLPLHLTLPLQALPYGA 292
DB 280 HASSGELEFIPNITTNNSGTYTCLVNNSVTGLSKRTTVK----- 316
QY 293 GSGNLTALAEAKTKLHOEVNLVVMRATOLQKNLTC-----EVMGPTSPKMLSLKEN 344
DB 317 ---NITV-LEPTVQPSLOVNTTVKELDSV--TLTCLSDNIGANIQWLFNSGSLQTERMTL 370
QY 345 ENKEAVSKREKRPVWVNLNPEAGMOCCLSDSGQVLLSEINIKVLPWTSTVPHRASALPAP 404
DB 371 TLSQNNSLIRIDP--IKREDAGEYOCEISNPVSVKRSNISKL-----DIIFD 415
QY 405 PTGSALPDPQTASALPDPPPASALPALA 433
DB 416 PTQGLSDGALIGIIVGAVGALTAGLA 444
```

```
RESULT 61
S34338
biliary glycoprotein F - mouse
N:Alternate names: mouse hepatitis virus (MHV) receptor glycoprotein
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 23-Jul-1999
C:Accession: S34338; JCI1510; A41093
R:Huang, D.C.; Huang, X.F.; Novel, M.; Novel, G.
submitted to the EMBL Data Library, July 1992
A:Description: A Clp-family gene present on the lactose-protease plasmid of Lactococcus
A:Reference number: S34338
A:Accession: S34338
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-521 <HNU>
A:Cross-references: EMBL:X67281; NID:g312585; PIDN:CAA47698.1; PID:g312586
R:McCuall, K.; Rosenberg, M.; Nedellec, P.; Turbide, C.; Beauchemin, N.
Gene 127, 173-183, 1993
A:Title: Expression of the Bgp gene and characterization of mouse colon biliary glycopro
A:Reference number: JCI1505; MUID:93273228; PMID:8500759
A:Accession: JCI1510
A:Molecule type: mRNA
A:Residues: 1-61, 'Q', 83-141, 'P', 143-521 <MCC>
A:Cross-references: GB:X67281
R:Williams, R.K.; Jiang, G.S.; Holmes, K.V.
Proc. Natl. Acad. Sci. U.S.A. 88, 5533-5536, 1991
A:Title: Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen
A:Reference number: A41093; MUID:91288498; PMID:1648219
A:Accession: A41093
A:Status: preliminary
A:Molecule type: protein
A:Residues: 35-59 <WIL>
C:Comment: This protein is expressed at the cell surface and plays a determinant role in
C:Genetics:
A:Gene: Bgpf
C:Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi
C:Keywords: glycoprotein; receptor
F:1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F:160-219/Domain: immunoglobulin homology <IMM1>
F:254-303/Domain: immunoglobulin homology <IMM2>
F:339-396/Domain: immunoglobulin homology <IMM3>
F:87,104,148,199,206,210,226,258,290,294,304,333,375/Binding site: carbohydrate (asn) (c
Query Match 5.0%; Score 116; DB 2; Length 521;
Best Local Similarity 19.7%; Pred. No. 3.5;
Matches 100; Conservative 74; Mismatches 189; Indels 144; Gaps 22;
QY 1 MNRG-VPRHLLLVLOL-----ALPPATQGNKVLGKKGDTVELTCTAS 44
DB 8 LHKGVPMVGGLLTSLASLWSPPTTAETIEAVPQVAEDNNVLLVHNLPALGA--- 64
QY 45 QKKSIOGFHWKNSNQIKILGNQGSFL-----TKRPSKLANDRADRSLSMDQGNPLIITKN 98
DB 65 -----FAMYKGNPVTNAEIVHVGVTGNTKTTPAH-----SGRETVYVSNGS--LLIQ 111
QY 99 LKIEDSDTYICEVEDQ-----KEEVQ-----LLVFGLTANSPTHLQOGSLTTLTESP 147
DB 112 VTVKDTGVYTIETMDENFRRTATVQFHNHPLLKPNITSNNSNPVEGDDSVSLTCDSTY 171
QY 148 GSSPVQCRSPRGKNI-----OGSKTLSVSOLELQDSGTWCTCTVLQNKQKVE--- 194
DB 172 DPNITVYLSMRNGBESLSEGRDLKLSBGNRTLLTNAVTRNDTGPVCEI--RNPVSVNRS 230
QY 195 FKIDIV-----VLAFOKASSIYKKKEGQEVESFPPLAFTVEKLTSGGELMW---QAE 245
DB 231 FSLNIIYGPDPPII-----SPSDIYLHPGSNLTNSCHAA-----SNPPAQYFWLINEXR 281
QY 246 SSSKSMI-----TFDLKNKEVSVKRVTPQDPKLGKGLPLHLTLPLQALPYGAS 294
DB 282 SSQELFIPNITTNNSGTYTCLVNNSVTGLSKRTTVK----- 316
QY 295 GNLTLALAEAKTKLHOEVNLVVMRATOLQKNLTC-----EVMGPTSPKMLSLKEN 346
DB 317 -NITV-LEPTVQPSLOVNTTVKELDSV--TLTCLSDNIGANIQWLFNSGSLQTERMTL 372
QY 347 KEAVSKREKRPVWVNLNPEAGMOCCLSDSGQVLLSEINIKVLPWTSTVPHRASALPAPT 406
DB 373 SQNNSLIRIDP--IKREDAGEYOCEISNPVSVKRSNISKL-----DIIFDPT 417
QY 407 GSALPDPQTASALPDPPPASALPALA 433
```

Db 418 OGGISDGAIAIGVIGVAGVALLIAGLA 444

RESULT 62

1JBOHC
neural cell adhesion molecule short domain form precursor - bovine
N:Alternate names: NCAM-140
C:Species: Bos primigenius taurus (cattle)
C:Date: 31-Mar-1993 #sequence revision 31-Mar-1993 #text_change 22-Jun-1999
C:Accession: A32976; A38778; B44290; S05402
R:Linkin, V.M.; Khramsov, N.V.; Andreeva, S.G.; Moshnyakov, M.V.; Petukhova, G.V.; Raki
FEBS Lett. 254, 69-73, 1989
A:Title: Calmodulin-independent bovine brain adenylate cyclase. Amino acid sequence and
A:Reference number: A32976; MUID:89378239; PMID:2776887
A:Accession: A32976
A:Molecule type: mRNA
A:Residues: 1-853 <LTP>
A:Cross-references: GB:X16451; NID:960; PIDN:CAA34470.1; PID:961
A:Accession: A38778
A:Molecule type: protein
A:Residues: 20-35;51-61;113-117;122-147;155-161;262-275;279-302;353-360;369-382;544-562;
A:Note: the authors identified this protein as calmodulin-independent adenylate cyclase
R:Rougon, G.; Marshak, D.R.
J. Biol. Chem. 261, 3396-3401, 1986
A:Title: Structural and immunological characterization of the amino-terminal domain of m
A:Reference number: A44290; MUID:86140120; PMID:3512556
A:Accession: B44290
A:Molecule type: protein
A:Residues: 20-36 <ROU>
A:Note: 23-Glu was also found
C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
C:Superfamily: Various forms of NCAM are produced by alternative splicing.
C:Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; s
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-853/Product: neural cell adhesion molecule, short domain form #status experimental
F:34-98/Domain: extracellular #status predicted <EXT>
F:132-191/Domain: immunoglobulin homology <IMM1>
F:152-166/Domain: immunoglobulin homology <IMM2>
F:161-165/Region: heparin binding #status predicted
F:228-288/Domain: immunoglobulin homology <IMM3>
F:261-270/Region: NCAM binding #status predicted
F:321-396/Domain: immunoglobulin homology <IMM4>
F:428-490/Domain: immunoglobulin homology <IMM5>
F:521-604/Domain: fibronectin type III repeat homology <FN3A>
F:631-693/Domain: fibronectin type III repeat homology <FN3B>
F:720-737/Domain: transmembrane #status predicted <TM>
F:738-853/Domain: intracellular #status predicted <INT>
F:41-96;139-189;235-286;328-394;435-488/Disulfide bonds: #status predicted
F:222;314;346;432;458;487/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.0%; Score 116; DB 1; Length 853;

Best Local Similarity 20.5%; Pred. No. 6.9;

Matches 87; Conservative 52; Mismatches 159; Indels 126; Gaps 18;

Qy 32 KKGDTVELTCTASQKKSIOFHMKNSNOIKIILNGSSFLTKGPKSLNDADSRRLMOGN 91
Db 130 REGDAVAVCDVSSLPRTIIMKHGRDVIILKQDRTIV-----LTN 171
Qy 92 PPLIINKLIEDSDTYICE-----VEDQKEVOLLV-----FGLTANSDPHL 133
Db 172 NYLQIRGRIKKTDEGYRREGRIILARGEINFDIQIVAVPPTVGAROSIVATAN----- 226
Qy 134 LOGOSLITLTPSSPSSSVQCRSPRKNIOG-----GKTLVSQLELDQSGTW 182
Db 227 -LGOSVITLVCAEGPEPTVSM-TKDGEIENEDEKTLFSDSSSLTIRKDKNDKDEAY 284
Qy 183 TCTVLONGKAYEFKIDIVLAFQKASIVYKKEGQVFSFPLAFTVVKLTGSGE----- 237
Db 285 VCIENKAGEODASIHLEKFAKPK--ITVENQGTAMELEQVITLTCE--ASGDPIPSI 338

Qy 238 LMMQERASSSK---SWTFDLKNKEVSRVTDPEKLOMGKLLPL-HLTLPQALPYAG 293
Db 339 TWRSTRNISSEKASWTRE-----KQETLDGMVVRSHARVSLTLKSI--QYTD 388
Qy 294 SGULTLALAKTGKLEHOEVLVVMRATOL-----QKNLTCEVMGPTSPKMLLS 341
Db 389 AGEVYCTASNTIQDSQSMYLEVOYAPKLOGPAVYVTEGNQVITCEVAPSP----- 442
Qy 342 LKLENKAKYSKREKPPVVLNPEAGMMQGLSDSGOVLLESNKVLPTWSPVHPASAL 401
Db 443 -----ATIS-----WFRD-----GQLPSSNYSNIKIYNT--PSASYL 473
Qy 402 PAPP 405
Db 474 EVTP 477

RESULT 63

139207
leukocyte surface protein V7 - human
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence revision 02-Jul-1996 #text_change 21-Jan-2000
C:Accession: I39207
R:Ruegg, C.L.; Rivar, A.; Madani, N.D.; Zeitung, J.; Laue, R.; Engleman, E.G.
J. Immunol. 154, 4434-4443, 1995
A:Title: V7, a novel leukocyte surface protein that participates in T cell activation.
A:Reference number: I39207; MUID:95238941; PMID:7722300
A:Accession: I39207
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1021 <RES>
A:Cross-references: EMBL:233642; NID:9854194; PID:9854195
C:Genetics:
A:Gene: GDB:V7
A:Cross-references: GDB:702141
A:Map position: IP13-IP13
C:Superfamily: human leukocyte surface protein V7

Query Match 5.0%; Score 116; DB 2; Length 1021;

Best Local Similarity 20.1%; Pred. No. 8.7; Indels 186; Gaps 27;

Matches 113; Conservative 69; Mismatches 195; Indels 186; Gaps 27;

Qy 16 LALLP--AATQGNKVLGKKGDVLELCTASQKKSIOFH-----KNSNQIKIL 62
Db 140 LIVIPDTLSATWSQTLGKEGEPLALTCASAKATAGHTLSVTVYLTQGGSGQATEIL 199
Qy 63 GNGQSF-LTGSPSKLNDADSRSLMDQ--NPPLIINKLIEDSDTYICE---VD-- 113
Db 200 SLSKDFLIVGPPYTERFAASDVQNLKGPTRFLSLIERLQSSDQGLFCEATEWIDPD 259
Qy 114 -----QKEVOLLV-----FGLTANSDPHLQGGSLTTL-LSSPPGSSPSVQ-- 154
Db 260 ETWMTFKKQTDQTLRIQPAVDFOVNITRADSIFAGKLELVCLVSSGRDQLOGIW 319
Qy 155 -CRSPRKNIOGKTLV-----SOLEIQ-----DSGTWTC- 185
Db 320 FNGTEIAHIDAGVGLKNDYKERASQGLQSLKGPFAFLKIFSLGPEDEGAYRCV 379
Qy 186 -----VLONQKAYEFKIDT-----VTLAFQKASIVYKKEGQVEF----- 221
Db 380 AEVWKRTGSGWOYLORQSPDSHVHLRKPARSVVSTKNKQOVW--BEETLAFICKAG 437
Qy 222 --SPPLAFT-----VEKLTGSGE-----LW-----MQERASSSKSWTFDLK 257
Db 438 GASSPLSVSMWHIIPRDOTQPEFVAGMGQDGIVOLGALLMGTSYHGNTRELKMDATQLE 497
Qy 258 -----NKEVSRVTDPEKLOMGKLL-----PLHLTPQALPYAGSGN 296
Db 498 ITFAITDSGYTCGRVSEKSRNQARDLSWQTKISVTVKSLSSLOVSLMSQPOVMLTNT 557
Qy 297 LTTALAEKTKGLHOEVLVVM-----RATLOKQLTCEVWGPTSPKMLLSKL 344
Db 558 FDLSCVVRAGYSDLKXPLVTVWQFOPASSHIFQLIRITHNGTIE--WGN-----FLSR 609

```
QY      345 ENKEAYKSKREKRVWVL-----NPEAGMGQCILSDSGQLVLESNTKYLPTMSTVFHHRAS   399
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      610 FOKTKVKVSGLPFSFSQLLVHDATGEQTGTVQCEVEVYDRNSLYNN-----        653
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      400 ALPAAPPGSALPPD-QTASALPD 421
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      654 ---PPPASAIISHPLRIAVTLPE 673
           ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 64
S57845
protein-tyrosine kinase (EC 2.7.1.112), receptor type tie precursor - bovine
C1:Species: Bos primigenius taurus (cattle)
C1:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 16-Jul-1999
C1:Accession: S57845; S32690
R1:Sato, T.N.; Qin, Y.; Kozak, C.A.; Audus, K.L.
Proc. Natl. Acad. Sci. U.S.A. 90, 9355-9358, 1993
A1>Title: tie-1 and tie-2 define another class of putative receptor tyrosine kinase genes
Reference number: S57845; MUID:94022374; PMID:8415706

A:Status: type:liminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-1136 <SMT>
A:Cross-references: EMBL:X71423, NID:q296575, PIDD:CA50554.1, PTD:q296576
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1993
C:Superfamily: protein-tyrosine kinase, receptor type tie; EGF homology; fibronectin type
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-116/Product: protein-tyrosine kinase, receptor type tie #status predicted <MAT>
F:36-108/Domain: immunoglobulin homology <IM1>
F:213-253/Domain: EGF homology <EG1>
F:257-300/Domain: EGF homology <EG2>
F:304-342/Domain: EGF homology <EG3>
F:363-426/Domain: immunoglobulin homology <IM2>
F:447-528/Domain: fibronectin type III repeat homology <FN3A>
F:540-631/Domain: fibronectin type III repeat homology <FN3B>
F:640-728/Domain: fibronectin type III repeat homology <FN3C>
F:759-784/Domain: fibronectin type III repeat homology <FN3D>
F:835-1112/Domain: protein kinase homology <KIN>
F:843-851/Region: protein kinase ATP-binding motif
F:843-106,370-424/Dissulfide bonds: #status predicted
F:84,159,501,594,707/Binding site: Cys, Asp, Asn (covalent) #status predicted
F:868,885,977/Active site: Lys, Glu, Asp #status predicted
F:868,885,977/Active site: Lys, Glu, Asp #status predicted

Query Match 5.0%; Score 115.5; DB 1; Length 1136;
 Best Local Similarity 21.4%; Pred. No. 11;
 Matches 86; Conservative 47; Mismatches 146; Indels 123; Gaps 16

Qy	109	CEVEDQKEEYQLLVFGITNSDPHLLQGSGLTLTESPPGSP-----SYGCRSPRKNI	163
Db	342	CEKSDRIPQLLDWSELEFNLDT-----MPRINCAAGIPFPFGSMELKRPDGTVL	393
Qy	164	QGGKTL-----SVSQLELDSDGTCTVLQN--QKVEFKIDIVLAFQKASSIV	211
Db	394	LSTKAIYEPDRTAEEFEVRPILAGDGLMECEVSTSGGDSRRFRINXKVPVPLTAPRL	453
Qy	212	YKKEGEVESPFLATFVEKLTGSGELTGMQARASSKSMITFDLKNKNEVVKRVTPDPK	271
Db	454	LAKSRQLVVSPLVSTSGGDIPLASVRLHYRPO--DSTAMST-----IVDP--	498
Qy	272	LQNGKKPLRLHLTPQALPOYAGSGNLTILA-LEAKTGKTLHQEYLVLMRATQLQKLTCEV	330
Db	499	-----SENVTLLKNLRLPKTG--YSVRQLSRPGEKGEG---A	529
Qy	331	MGPTS-----PKTMTSLKLENKEAVKSGREKPVWVNLPEAG-----MMQCLISD	374
Db	530	MGFPTLMTTCPEPFLKPKMLEGHNHVGPRDLKAVNSLSPVPQPLVGDFGLRLMD---GA	586
Qy	375	SGGVLLLESNIK-----VLPTWST-----PVHPRASALPAPPTGSAL	410
Db	587	RGGERRR-NVSSPOARTALLTGLTPEYVQLDVRLYHCTGLLGPASPAARVLI-LPPSGPFA	644

```

Oy      411 PDPQTASALPDP-----PASAALPALAVISFLGLG 444
          ||| |
Db      645 PRLHAQALSDSEIQLMQRPEAAGPISKYIVEVQVAGSG 686

```

RESULT 65
 A69332
 hypothetical protein AF1138 - Archaeoglobus fulgidus
 C/Species: Archaeoglobus fulgidus
 C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
 C/Accession: A69332
 R/Klink, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 .; Gloeck, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 . Nature 390, 364-370, 1997
 A/Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artchik, P.; Kaine, B.P.; Sykes, S.
 . Smith, H.O.; Woese, C.R.; Venter, J.C.
 A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
 .
 A/Reference number: A69250; PMID:98049343; PMID:93899475

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-554 <KLE>
A:Cross-references: GB:AE001025; GB:AE000782; NID:g2669348; PID:g264945
A:Residues: 1-554 <KLE>

Query Match	4.9%	Score 115;	DB 2;	Length 554;
Best Local Similarity	20.6%	Pred. No. 4.5;		
Matches 119; Conservative	77;	Mismatches 204;	Indels 178;	Gaps 27

```

Qy      7 FRHLVLVLALLPATGCKNVILGKGDVETLTCTASQKSIQFHWKNSNQIKILNQ 66
          || || : : : || || :
Db      4 FRFLLAATALLTGSIE-----CELKEITAPGEEVISPLYKND-----GNES 49

```

```

Qy      67 SFL-----TKGPSKLNDRADSRSLWDQGNFPIIKUKIEDSDTYI----- 108
      : | | | | | | | | | |
Db      50 TFOLSYSFWYGEAEGFYFNQGRVNSLRLNSSSESADITFKFLAPEKGRGYYLYLHADSSA 109

```

```
Qy      109 -----CEVEDQKEVQLLVGELTANSSTHLIQGSILTTLESPPGSSPVSQCSPRGKNI 163
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      110 GVTAVNELPEKPLRISLDNTGIVAEGSDVEVNALLIENTLNSP--IEVDLSCAAPKMGNEC 167
```

```

Oy 164 -----QGGKLTLSVSQLELQ---DSGWTCTYLQNLQNKVEFKIDIVLAFQ 205
      :||| | |:: | | | |:: |
Db 168 RFYDGDVEYRTVVEGGGKQLRV-QVDIDSTADVGKAVTLIHGPQKEFEV-----FV 220

```

```

Oy      206 KASSI-----VYKKEGHQVSEFSPLA-----PTVEKLTGSGELIMWQAERASSKSW 251
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      221 KESHAGEKGEVRLRVVDKDGEGVASAKRITAGNETFT-----SGDGEAIFEEVCG----- 270

```

```
QY      252 ITFDIK-NKEVSVKRVTTQDPKLQMGGKLPHLTLPALPOYA-----292
          ||||| :: :| :| :| :| :| :|
DB      271 -TYDLEIEKGGYEKTIRDVAVRGGRPTNDLCTVFLEKAYIAELSVSSSRITWISIGDYAS 322
```

```
Qy      293 -----GSGNLTALAA-----KTGKHQEVNLVVMRATOLQKULTCEWGP 333
          |||           :|||   |   :   :|||   |||
Db      330 LNVRIENRGYGEDQYALSTGLPPPTYNREGEL--AVSEFYLEGGE-EKITLLEIYTP 366
```

```

Oy      334  TSPKL-MLSKL-----ENKEAKYSKREKPVMTLNPAGMQCLLS----- 373
      : :: ||| : : : : : : : : : : : : : : : : : : : : : :
Db      387  ATAELGDISLKIIVAKGYTAEKDLKLSVVGMAAYISLEGMYTTVKASPGEELELKGVE 446

```

```
Qy      374 DSGVLLSENIKV---LPT-WSTPHPRASALPAPPTGSLPDPQTASALPPPPAASALP 429
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      447 NSGRVTLTIVKVKLELPQGMSGEVSP--EVIPSVKRGER---GEVNLRIRVPDAKPE 500
```

Qy 430 AALAV-----ISFLG-----LGLVACV 448
 : | : : : :
 Db 502 YRIIVETTSDDMTMSDRISVVVGESSATFVGLGIYAV 519

RESULT 66
I38344
titin, cardiac muscle [validated] - human

N:Alternate names: connectin
 N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.1-)
 C:Species: Homo sapiens (man)
 C/Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text_change 15-Sep-2000
 C/Accession: 138344; S20897; S20899; S63665; S37393
 R/Label: S.; Kolmerer, B.
 Science 270, 293-296, 1995
 A>Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
 A:Reference number: A57430; MUID:96026330; PMID:7569978
 A:Accession: 138344
 A>Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL
 A:Molecule type: mRNA
 A:Residues: 1-26926 <LAB>
 A:Cross-references: EMBL:X90568; NID:g1017424; PID:g1017425
 R/Musco. G.; Tiazzi, C.; Schuck, P.; Pastore, A.
 Biochemistry 34, 553-561, 1995
 A>Title: Dissecting titin into its structural motifs: identification of an alpha-helix
 A:Reference number: 138345; MUID:95119041; PMID:7819249
 A:Accession: 138345
 A>Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1377-2014 <MUS>
 A:Cross-references: EMBL:X83270; NID:9602579; PIDN:CAA58243.1; PID:9602580
 A>Note: Conformation and properties are reported for a synthetic peptide corresponding to
 R/Label: S.; Gautel, M.; Lahey, A.; Trinick, J.
 EMBO J. 11, 1711-1716, 1992
 A>Title: Towards a molecular understanding of titin.
 A:Reference number: S20897; MUID:92258380; PMID:1582406
 A:Accession: S20898
 A>Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 13597-14200; 'I', 14202-14696 <LAB2>
 A:Cross-references: EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID:g37193
 A:Accession: S20897
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 16330-16362; 'S', 16384-16756; 'F', 16758-16860 <LAB3>
 A:Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191
 A:Accession: S20899
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 'P', 22278-22431; 'R', 22433-22448; 'G', 22450-22453; 'Q', 22455-22480; 'TR', 22483-2
 A:Cross-references: EMBL:X64697; NID:g37190; PIDN:CAA45938.1; PID:g37195
 R/Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labelle, S.
 J. Mol. Biol. 256, 556-563, 1996
 A>Title: Genomic organization of M line titin and its tissue-specific expression in two
 A:Reference number: S63665; MUID:96177761; PMID:8604138
 A:Accession: S63665
 A>Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 26729-26825 <COL>
 A:Cross-references: EMBL:X92412; NID:g1236761
 R/Gautel, M.; Leonard, K.; Labelle, S.
 EMBO J. 12, 3827-3834, 1993
 A>Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentiat
 A:Reference number: S37393; MUID:94008990; PMID:8404852
 A:Accession: S37393
 A:Molecule type: mRNA
 A:Residues: 26831-26926 <GAU>
 R/Improta, S.; Politou, A.S.; Pastore, A.
 submitted to the Brookhaven Protein Data Bank, February 1996
 A:Reference number: A66736; PDB:1TTT
 A:Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
 R/Futtl, M.; Pastore, A.
 submitted to the Brookhaven Protein Data Bank, August 1996
 A:Reference number: A66201; PDB:1NCT
 A:Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155
 C/Genetics:
 A:Gene: GDB:TTN
 A:Cross-references: GDB:127867; OMIM:188840
 A:Map position: 2q31-2q32
 C/Function:
 A>Description: structural protein forming filaments in striated muscle

C/Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; prot
 C/Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco
 structural protein
 F:24752-25008/Domain: protein kinase homology <KIN>
 F:184,177,905,2276,2378,2561,2569,2669,2763,2896,3088,3179,3384,3432,3628,3772,4068,4
 98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1
 16780,16976,17579,17602,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680,186
 21900,21935,22295,22495,22627,22827,23024,23318,23883,24012,24177,24290,24447,24642,248
 F:26171,26178,26184,26190/binding site: phosphate (Ser) (covalent) #status experimental
 Query Match 4.3%; Score 115; DB 1; Length 26926;
 Best Local Similarity 19.3%; Pred. No. 8.3e+02;
 Matches 115; Conservative 73; Mismatches 184; Indels 224; Gaps 26;
 QY 11 LVLQALALPAA-TOG--KNVYLGKGDVYELCTASQKKSIGFHMNSQIKLNGQS 67
 DB 14557 VIVEQTMLELDLRGYYQLVIAKAGDNKIKVPIVYGRPKPTVYWKQQLKQTRVN 14616
 QY 68 FLTKGPSKLNDRADSRSLWDQGNFPIIKNLKTEDSDTYCEVED-----QKEEV 118
 DB 14617 FETTAISTININECVRS--DSGPYPLTANIVGEVDVTTIQVHDIPGPTGPIKDEV 14674
 QY 119 QLLVFGLTANSDFHLLOGSLVTLTLESPPSSSPVQCRSPRGKNIQGKTLG--VSQLEL 176
 DB 14675 -----SSD-----FVTFSDMPPEV-----DGVPISNVYVEMRQ 14703
 QY 177 QDSGTW---TCYVLQNKRV-----EFKID-----IV----- 200
 DB 14704 TDSITWELATTVIRTYKATRLTLTGLEVQFRVKAQNRGVGPGITSAMIVANYPFRVPG 14763
 QY 201 -----VLAFQKASIVYKKEGQVEFSPLATFVEKLTGSGELMQAERA----- 245
 DB 14764 PRGPQVTAATKOSMTISWHEPLSDGSPRLTGHNVERKENGLMQVSKALVPGNIFKS 14823
 QY 246 SSSKSWITFDL-----KNKEVSVKRVTDPRKLQNGKKLPHL-----TLQALP 289
 DB 14824 SGLTDGIAVFRVIAENMAKSKRSPSEPMALDIPDPGKVPINITHVTWLKWAAP 14883
 QY 290 QVAGSGULTALAEK---TGKHHQ-----EVNLVVMATQ-----LQKNLTCEV 330
 DB 14884 EYTGGEFKITIVYEKDLDPNGRWLKNFNSILNEFTSGLTDAVAEFVIAKNAAGAI 14943
 QY 331 MGPT-----SPKMLSLKLNK-----EAKVSREKRPV--W----- 359
 DB 14944 SPSESEBDATTCDDVDVAPKIKVDVAKFQVILKAGAFPLEADVSRPPTMEKSDGK 15003
 QY 360 -----VLPBAGMWQCLSDSGQVLES-----NIKVLPTW 390
 DB 15004 ELEGTALEIKIADFTSTNLVVKDS-----TRDSGAYTLTATNPGFAKHIFNVKVLDR- 15057
 QY 391 STVPHRASALRPAPGASALPDPQTASALPDPRAASLPAALVISPILGLIGLVA 446
 DB 15058 -----PQPPG-----PLAVTEVISEKCVLSWFPPLDDGA 15088
 RESULT 67
 S00682
 IGE Fc receptor alpha chain precursor - human
 N:Alternate names: Fc-epsilon receptor
 C/Species: Homo sapiens (man)
 C/Date: 31-Dec-1988 #sequence revision 31-Dec-1988 #text_change 23-Jul-1999
 C/Accession: S00682; B30154; S42209
 R/Kochan, J.; Pettine, L.F.; Hakim, J.; Kishi, K.; Kinet, J.P.
 Nucleic Acids Res. 16, 3584, 1988
 A>Title: Isolation of the gene coding for the alpha subunit of the human high affinity Ig
 A:Reference number: S00682; MUID:88233953; PMID:2967464
 A:Accession: S00682
 A:Molecule type: mRNA
 A:Residues: 1-257 <KOC>
 A:Cross-references: EMBL:X06948; NID:g31317; PIDN:CAA30025.1; PID:g31318
 R/Shimizu, A.; Teppler, I.; Bentley, P.N.; Berenstein, E.H.; Stragatian, R.P.; Leder, P.
 Proc. Natl. Acad. Sci. U.S.A. 85, 1907-1911, 1988

A>Title: Human and rat mast cell high-affinity immunoglobulin E receptors: characterization
A:Reference number: J94191; MUID:80150102; PMID:2964640
A:Accession: B30154
A:Molecule type: mRNA
A:Residues: 1-257 <SH>
A:Cross-references: GB:J03605; NID:9187449; PID:AAA36204.1; PID:9307164
R:Yagci, S.; Yanagida, M.; Tanida, I.; Hasegawa, A.; Okumura, K.; Ra, C.
Eur J. Biochem 220, 593-598, 1994
A>Title: High-level expression of the truncated alpha chain of human high-affinity receptor product.
A:Reference number: S42209; MUID:94170811; PMID:8125119
A:Accession: S42209
A:Molecule type: protein
A:Residues: 26-197 <YAG>
A:Experimental source: purified recombinant protein
C:Genetics:
A:Gene: GDB:FCER1A
A:Cross-references: GDB:119902; OMIM:147140
A:Map position: 1q23-1q23
C:Superfamily: Fc gamma receptor II; immunoglobulin homology
C:Keywords: immunoglobulin receptor; transmembrane protein
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-257/Product: 198 Fc receptor alpha chain #status predicted <MAT>
F:44-95/Domain: immunoglobulin homology <IMM1>
F:125-178/Domain: immunoglobulin homology <IMM2>

Query Match 4.9%; Score 114; DB 2; Length 257;
Best Local Similarity 26.2%; Pred. No. 1.9;
Matches 73; Conservative 34; Mismatches 92; Indels 80; Gaps 15;

QY 10 LLLVQLALP-----AATGKNVVGK-----KGDTELTCTAOKKSIOFHMKSNQI 59
DB 10 LLLVQLALPAPGVLAVPQKPVNSLPNNRKFKEBNTLTC----- 51
QY 60 KILGQSEFLTKPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEV--DQKEE 117
DB 52 ----NGNNFEVSSRTKMFNG--SLSEETNSLNTVNAKEDSGEYKCHQGVNEBP 103
QY 118 VQLVFGLTANSDFHLQGSITLTLSPGSSPSVQCRSPGKNI-----QSGKITL-- 169
DB 104 VLEVF-----SDWLLQ--SAEVMVEGP--LFLRGCHGMNMDVYKVIYKDEALKY 154
QY 170 -----SVSQLEDDSGTWCT--VLNQKVEFKIDIVLAFQKASIVYKKEGQVE 220
DB 155 WENNNISTNATVDSGYTGTGKVMQDYDS--PLNTVY--KA-----PREKYLQ 205
QY 221 FSPPLAFVTEKLTGSGELMWQAEASSSKSWITFDLKNK 259
DB 206 FPIPLVILFAVDGLF-----ISTQQVTFILKIK 237

RESULT 68
C42632
cell adhesion molecule apCAM (clone d12) - California sea hare
C:Species: Aplysia californica (California sea hare)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 17-Mar-2000
C:Accession: C42632
R:Mayford, M.; Barzilai, A.; Keller, F.; Schacher, S.; Kandel, E.R.
Science 256, 638-644, 1992
A>Title: Modulation of an NCAM-related adhesion molecule with long-term synaptic plasticity.
A:Reference number: A42632; MUID:92263095; PMID:1585176
A:Accession: C42632
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-765 <MAX>
A:Experimental source: CNS
C:Note: sequence extracted from NCBI backbone (NCBIP:101351)
C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immunoglobulin

Query Match 4.9%; Score 114; DB 2; Length 765;
Best Local Similarity 18.1%; Pred. No. 8;
Matches 65; Conservative 59; Mismatches 119; Indels 116; Gaps 17;

QY 29 VLGGKQDVELTCTASOKKSIOFHMKSNQIKILGQSFLLTKGPKLNDRADSRSLMD 88
DB 136 ILGSG--EVECEVSGKPATVYWKPENNKIDAG-----KTTALN----- 176
QY 89 QGNFPLIKNLKIEDSDTYICEV-----EDQKEVOLLVFGL-----TANSDTHLQ 135
DB 177 ----KLITKDSLTDKTKYLCDIIVDPTGETKDYIFPTVVKLPITALPPIHDPNKV- 231
QY 136 QGSITLTLSPGSSPSVQCRS----PRGNIOGKTLVSQLELDSGTWCTV----- 186
DB 232 GDEKVTICQAGVPPYQFKKGDMVTDEWVNNG-VLTINPLKTTQATYTTCATNKG 290
QY 187 -LONOKVEFKI-----DIVLAFOKAS--SIYKKEGQVEFSPFL 225
DB 291 FAESSNLDVVPPIETMEETVDAVSGELITTCATKDEPSPVINKDG-----PQ 343
QY 226 AFTVEKLTGSGELMWQAEASSSKSWITFDLKNKESVYKRVTDPKLQMGKKLPLHLTP 285
DB 344 SASDGIYVKGPTV--EKVGSQN--DWEKTVV-----QHMTEK 379
QY 286 QALQYAGSGNLTALAEKTKGLHGEVNLVYMRATQ-----QKNLTGE 329
DB 380 PV--TYDAGTYICTASFLVGSANKTVLTVQYKPNPDTEKEREFGMRGKANTCQ 436

RESULT 69
C42632
cell adhesion molecule apCAM (clone d15) - California sea hare
C:Species: Aplysia californica (California sea hare)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 17-Mar-2000
C:Accession: B42632
R:Mayford, M.; Barzilai, A.; Keller, F.; Schacher, S.; Kandel, E.R.
Science 256, 638-644, 1992
A>Title: Modulation of an NCAM-related adhesion molecule with long-term synaptic plasticity.
A:Reference number: A42632; MUID:92263095; PMID:1585176
A:Accession: B42632
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-812 <MAX>
A:Experimental source: CNS
C:Note: sequence extracted from NCBI backbone (NCBIP:101346)
C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immunoglobulin

Query Match 4.9%; Score 114; DB 2; Length 812;
Best Local Similarity 18.1%; Pred. No. 8.7; Indels 116; Gaps 17;
Matches 65; Conservative 59; Mismatches 119; Indels 116; Gaps 17;

QY 29 VLGGKQDVELTCTASOKKSIOFHMKSNQIKILGQSFLLTKGPKLNDRADSRSLMD 88
DB 136 ILGSG--EVECEVSGKPATVYWKPENNKIDAG-----KTTALN----- 176
QY 89 QGNFPLIKNLKIEDSDTYICEV-----EDQKEVOLLVFGL-----TANSDTHLQ 135
DB 177 ----KLITKDSLTDKTKYLCDIIVDPTGETKDYIFPTVVKLPITALPPIHDPNKV- 231
QY 136 QGSITLTLSPGSSPSVQCRS----PRGNIOGKTLVSQLELDSGTWCTV----- 186
DB 232 GDEKVTICQAGVPPYQFKKGDMVTDEWVNNG-VLTINPLKTTQATYTTCATNKG 290
QY 187 -LONOKVEFKI-----DIVLAFOKAS--SIYKKEGQVEFSPFL 225
DB 291 FAESSNLDVVPPIETMEETVDAVSGELITTCATKDEPSPVINKDG-----PQ 343
QY 226 AFTVEKLTGSGELMWQAEASSSKSWITFDLKNKESVYKRVTDPKLQMGKKLPLHLTP 285
DB 344 SASDGIYVKGPTV--EKVGSQN--DWEKTVV-----QHMTEK 379
QY 286 QALQYAGSGNLTALAEKTKGLHGEVNLVYMRATQ-----QKNLTGE 329
DB 380 PV--TYDAGTYICTASFLVGSANKTVLTVQYKPNPDTEKEREFGMRGKANTCQ 436

RESULT 70

A42632
 cell adhesion molecule apCAM (clone d19) - California sea hare
 A:Cross-references: GB:M21622; NID:g204109; PIDN:AAA41146.1; PID:g204110; GB:J03811
 C:Species: Aplysia californica (California sea hare)
 C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 17-Mar-2000
 C:Accession: A42632
 R:Mayford, M.; Barzilai, A.; Keller, F.; Schacher, S.; Kandel, E.R.
 Science 256, 638-644, 1992
 A:Title: Modulation of an NCAM-related adhesion molecule with long-term synaptic plasticity
 A:Reference number: A42632; MUID:92263095; PMID:1585176
 A:Accession: A42632
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-932 <MAX>
 A:Experimental source: CNS
 A>Note: sequence extracted from NCBI backbone (NCBI:101342)
 C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu

Query Match 4.9%; Score 114; DB 2; Length 932;
 Best local similarity 18.1%; Pred. No. 10;
 Matches 65; Conservative 59; Mismatches 119; Indels 116; Gaps 17;

Oy 29 VLGGKGVDELCTRSQKSIQFHKNKNQIKLGNQSFILTKGSKLNDRAADRSRLMD 88
 Db 136 ILGGEG---EVECEVSGRPAPVTWKFENNTKIEAG---KXTIALN----- 176
 Oy 89 QGNFPLIKNKIEPDSPTVCEV-----EDQKEVQLVFGF-----TANSDTHLQ 135
 Db 177 ----KLIIKDLSTBETKRYKLCIDIVDGETKDFIDPTVKLPRIALPHIHNDNRY- 231
 Oy 136 GQSLTLTLESPGSSPSVQCRS---PRGNIOGGKTLVSQLELDGSGTWCTV----- 186
 Db 232 GDEVITCGATGVPPPTVQFKKGDMVTDENWNG-VLTINPLKTTDQATYTCIATNKG 290
 Oy 187 -LQNGKKEFKI-----DIVVLAFOQAS---SIVYKKEEQVEFSPRL 225
 Db 291 FAESNTLDVAVPPTIEDMEETVDVSGOELTITCTAGDPEPSYIMWKDQ-----PQ 343
 Oy 226 AFTVEKLTGSELMMQAEARASSKSWITFDLKNKEVSKVATODPKLOMGKKLPLHLTLF 285
 Db 344 SASDGIYNNKPTV---EKVSGNQN---DMEKTKVA-----QHMTFK 379
 Oy 286 QALPOYAGSNULTALEAKTGKHOENVLVVNRATQL-----QKNLTCE 329
 Db 380 PV--TYQAGTYITCTAFSLVGSANKTKLTVQYKPNFDTDEKEREFFGMRGKANKLTQ 436

RESULT 71
 A30154
 19B receptor alpha chain precursor - rat
 N:Alternate names: Fc-epsilon-R alpha chain precursor
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 18-Oct-1989 #sequence_revision 18-Oct-1989 #text_change 21-Jan-2000
 C:Accession: C31327; A31327; A30154; A27116; I55304
 R:Lin, F.T.; Albrandt, K.; Robertson, M.W.
 Proc. Natl. Acad. Sci. U.S.A. 85, 5639-5643, 1988
 A:Title: cDNA heterogeneity suggests structural variants related to the high-affinity Ig
 A:Reference number: A44203; MUID:88289772; PMID:2969594
 A:Accession: C31327
 A:Molecule type: mRNA
 A:Residues: 1-245 <L13>
 A:Cross-references: GB:M21622; GB:J03811
 A:Experimental source: basophilic leukemia cell line, clone R3-4
 A:Accession: A31327
 A:Molecule type: mRNA
 A:Residues: 21-245 <L13>
 A:Cross-references: GB:M21622; NID:g204109; PIDN:AAA41146.1; PID:g204110; GB:J03811
 A:Experimental source: basophilic leukemia cell line
 R:Shimizu, A.; Teppler, I.; Benfey, P.N.; Berenstein, E.H.; Stragmanian, R.P.; Leder, P.
 Proc. Natl. Acad. Sci. U.S.A. 85, 1907-1911, 1988
 A:Title: Human and rat mast cell high-affinity immunoglobulin E receptors: characterization
 A:Reference number: A34191; MUID:86158102; PMID:2964640
 A:Accession: A30154
 A:Molecule type: mRNA

A:Residues: 1-245 <SH1>
 A:Cross-references: GB:J03606; NID:g205331; PIDN:AAA41582.1; PID:g205332
 R:Kinet, J.P.; Metzger, H.; Hakim, J.; Kochan, J.
 Biochemistry 26, 4605-4610, 1987
 A:Title: A cDNA presumably coding for the alpha subunit of the receptor with high affinity
 A:Reference number: A27116; MUID:88024987; PMID:2959318
 A:Accession: A27116
 A:Molecule type: mRNA
 A:Residues: 1-'G', 3-236, 'N', 238-244, 'RLKPNK' <KIN>
 R:Teppler, I.; Shimizu, A.; Leder, P.
 U. Biol. Chem. 264, 5912-5915, 1989
 A:Title: The gene for the rat mast cell high affinity IgE receptor alpha chain. Structure
 A:Reference number: I55304; MUID:89174653; PMID:2522441
 A:Accession: I55304
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-16 <RBS>
 A:Cross-references: GB:M25334; NID:g341335; PIDN:AAA74562.1; PID:g556391
 C:Superfamily: Fc gamma receptor II; immunoglobulin homology
 C:Keywords: immunoglobulin receptor; transmembrane protein
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-245/Product: IgE receptor alpha chain #status predicted <MAT>
 F:42-93/Domain: immunoglobulin homology <IMM>

Query Match 4.9%; Score 113.5; DB 2; Length 245;
 Best local similarity 23.5%; Pred. No. 1.9;
 Matches 63; Conservative 41; Mismatches 85; Indels 79; Gaps 14;

Oy 11 LVLQALPLPAATQGNKTVLGK-----KQDTVELTC---TASQKSIQFHKNKNQIKI 61
 Db 13 LVLISLQVMTLATQKSVSLDPPMIRLTGDKTLLCNGNNSQMMSTK--WINDSI-- 68
 Oy 62 LGNGSFLTKGPKLNDRAADRSRLMDQGNFPLIKNLKIEPDSPTVCEVED--QKEVQ 119
 Db 69 -----SNVKSNNH-----VYSATIQDSGKTIQKQGFYKSPRY 103
 Oy 120 LLYVGLTANSDTHLQGSQSLTLTLESPGSSPSVQCRSPRG-----KNIQ 164
 Db 104 LNNV-----QEWLLQLQSSADVLDN---GSFDIRGSGMKWVKNVYKYKDIAFYKSY 154
 Oy 165 GSKTLSVQLELDGSGTWCTTVLQNGKKEFKLIDIVLAFOKASIVYKKEGEVEFSFP 224
 Db 155 DSNNISIRKATFNDSGSYHCTGYLN--KYCKSKDKESIAVVKDYTYEYR---WLQILF 208
 Oy 225 ----LAFTVEKLTGSELMMQAEARASSS 248
 Db 209 SLAVILPAVD--TG---LWFSTHKQFES 231

RESULT 72
 JC5519
 50K glycoprotein precursor - chicken
 C:Species: Gallus gallus (chicken)
 C>Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 17-Nov-2000
 C:Accession: JC5519
 R:Hancock, K.A.; Gooley, A.A.; Jeffrey, P.L.
 Mol. Brain Res. 44, 273-285, 1997
 A:Title: Acgp50, a predominantly axonally expressed glycoprotein, is a member of the Igk
 A:Reference number: JC5519; MUID:97225899; PMID:9073169
 A:Accession: JC5519
 A:Molecule type: mRNA
 A:Residues: 1-338 <HAN>
 A:Experimental source: brain
 C:Comment: This protein belongs to the IgLON's subfamily of cell adhesion molecules.
 C:Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
 C:Keywords: glycoprotein
 F:1-31/Domain: signal sequence #status predicted <SIG>
 F:32-338/Product: 50K glycoprotein #status predicted <MAT>
 F:46-113/Domain: immunoglobulin homology <IMM>
 F:146-199/Domain: immunoglobulin homology <IMM2>
 F:232-282/Domain: immunoglobulin homology <IMM3>
 F:40,136,148,279,300,315/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 4.9%; Score 113.5; DB 2; Length 338;
Best Local Similarity 21.3%; Pred. No. 2.9; Indels 101; Gaps 20;
Matches 84; Conservative 63; Mismatches 146; Indels 101; Gaps 20;

10 LLLVQLALLPA-----TQGNKVVLGKKGDTVELTCTASOKKSIQFMKNSNOIKI 61
14 LVLRLCLLPGLPVRSDFTGRGDNITVAGDFAILRCFVEDRSS-KVAMLRNSGIIF 72
62 LGNOSFLTGPBKSINDPAD-SRSLMDQGNFPLIINKLKIEDSTTYICEVEDQ---KE 116
73 AG-----EDKMSIDPVELEKRSPLR---YSLRIQVDYDESSYCSYQTOQHHPKTS 122
117 EVOLLVFG-----LTANSPTHLLQGSLTLTLESPPGSSPSVOCR--SPRKNIOGGKT-L 169
123 QVYLIVQVPKTSINSSDITTVESGNTLVCMANGRPPPVITWRHLITPGKEFBGEEYL 182
170 SVSOLLEODSGTWCTVQLONOKVFKIDIVLAFOKASSIYVKEGEOVE--SFPLPAF 227
183 ELIGTRGSGKYEC-----KANEVASADVAVQVAVTVVYPTI 221
228 TVEK-----LTSGELMWQAEERASSKSWITPDLKNKEVSVKRVITODPKLQMGKLPPLHT 283
222 TESKSNBATGR-----QALLPCEASAVPTPDF-----EWRDTRINSANGLEIKST 269
284 LPQALPOYAGSGNLTALAEAKTGKLGQEVNLVWMRATOLQKNLTCEVWGPTSPKLM--- 340
270 -----GSGSLMWANVTEHYG-NITCVANAL-----GVNNSLYLYRP 308
341 -SLKLENKAKVSKREKPVWVLPNPAAGMOCILS 373
309 GTGRVDNGSVSLA---VPLMLL---AASLCLLS 336

RESULT 73
T42405
sax-3 protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C:Accession: T42405
Rizalien, J.A.; Yi, B.A.; Bargmann, C.I.
Cell 92, 217-227, 1998
A:Title: The conserved immunoglobulin superfamily member SAX-3/Robo directs multiple asp
A:Reference number: 222160; MUID:98117250; PMID:9458046
A:Accession: T42405
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1273 <ZAL>
A:Cross-references: EMBL:AF041053; NID:G2804779; PIDN:AAC38848.1; PID:G2804780
C:Genetics:
A>Note: SAX-3
C:Function:
A:Description: sax-3 function is required at the time of axon guidance

Query Match 4.9%; Score 113.5; DB 2; Length 1273;
Best Local Similarity 19.9%; Pred. No. 17; Indels 193; Gaps 24;
Matches 106; Conservative 59; Mismatches 175; Indels 193; Gaps 24;

2 NRGVFRHLVLVQLA-----LLPATQGNKVVLGKKGDTVELTCTASOKKSIQF 51
3 NRKTLCTLLVLQAVIRSFCEASGNLAPVIEHDIVVSRGSPATLNCGA-KPSTAKI 61
52 HMKNSNQIKILGNOSFLTGPBKSINDPADSRSLMDQGNFPLIIT---KNLKIEDSTY 107
62 TWYKDGQ--PVITNK-----EYNSHRIYLDIGSLFLLVKNSGKNGKSDAGAY 108
108 ICEVEDQKEVOLLVFGLTANSPTHLLQGS/LTL-----TESPPGSSPSVQC 155
109 YCVANENGEVYK-----SNEG-----SLKLMLEDRPRVRRRTVQALGEMAVLEC 154
156 RSPRG-----KNIOGGKTLVSQLELQDSGTWCTVQLONOKV 193
155 SPFRGFPFPPVSVWRKDKELRIQDMPRYTLHSDGNLIIIDPVDRSDSGTYGCV----- 206

194 EFKIDIVLAFOKASSIYVKEGEOVEFSPLAFTVEKLTGSGELMWQAEERASSKSWIT 253
207 -----ANNV-----GERV--SNPRLSV-----FEKRFEBEPDMT 237
254 FDLKNKEVSVKRVITODPKLQW---GKKLPLHLTLPQALPOYAGSGNLTALAE----- 303
238 VDVGAVALPDCRVGDPOQPTWKRKNEPMPVT-----RAVIADKNGLRIERVQPSDEG 292
304 -----KTKGKHOEVNLVWMRATOLQ-----KNLTCEVWGPTSPK----- 337
293 EYVCYANPAGTLEASHLRVQAPPSFQTRPADQSVAGGTATFECTLVQQSPAYWMSK 352
338 -----LMLSLKENKAKVSKREKPVWVLPNPE-----AGMOCIL-LSDSGVLLSNI 384
353 EGQODLLFPSTVASDGTQTS---PYGTLIEVRQVDEGAYCAGMNSAGSLSTAAL 408
385 KVLDP---TWSTPVHPRASALPAPT-----GSALPPOQTASALPDP 422
409 KVTTKAVTGNTPAKP-----PPTIEHGQNGTLMWGSSAILPQASGKPTP 454

RESULT 74
T30532
neural cell adhesion molecule L1 homolog - Fugu rubripes
C:Species: Fugu rubripes
C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 01-Mar-2002
C:Accession: T30532
R.Riboldi Tumnicliffe, G.R.; Platzer, M.; Nakamura, G.; Elgar, G.S.; Brenner, S.; Rosen
submitted to the EMBL Data Library, September 1997
A:Description: Analysis of the genomic loci of Fugu rubripes homologs of the human disease
A:Reference number: 220848
A:Accession: T30532
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1277 <RIB>
A:Cross-references: EMBL:AF026198; NID:G3098263; PID:G3098264; PIDN:AAC15880.1
C:Genetics:
A:Introns: 42/1; 47/1; 81/2; 149/1; 190/1; 247/1; 285/2; 347/1; 391/1; 440/1; 477/2; 531/
2
A>Note: L1-CAM
C:Superfamily: neural cell adhesion molecule L1, fibronectin type III repeat homology; I

Query Match 4.9%; Score 113.5; DB 2; Length 1277;
Best Local Similarity 21.7%; Pred. No. 17; Indels 81; Gaps 16;
Matches 80; Conservative 56; Mismatches 152; Indels 81; Gaps 16;

34 GDTVELTCTASOKKSIQFMKNSNOIKILGNOSFLTGPBKSINDPADSRSLMDQGNFP 93
363 GETVKLDCQADGIRSPITW-----TVNGVPLSATSLSPRRSLTESGS-- 405
94 LIINKLKIEDSDTYICEVEDQKE-----VQLLVFGLTANSPTH-LIQGS/LTL-- 141
406 LILKDVIFGDTALIQCASNKHGTILANTWVYIELPPOILTEGNTYTFEGGKALLEC 465
142 -TLESP-----PSSSVQCRSPRKNIOGGKTLVSQLELQDSGTWCTVQLONOKVE 194
466 ETPGSPKPKTWESSSLIADPRVNLITNG-GLEIANSYHDEBGIYITCLVQGSNISVN 524
195 FKIDI-----VLAFOKASSIYVKEGEOVEFSPLAFTVEKLTGSGELMWQAEER---ASS 247
525 AEVEVNLNRVILSPQALRL---QPKTAIFTC-LVYTDPKLSPLQMKNDKXIFESH 580
248 SKSWITFDLKNKEVSVKRVITODPKLQMGKLPPLHTLPQALPOYAGSGNLTALAEATGK 307
581 SDKRYTFDGPGLIIS-----NVEPDGEGVYTCQIITKIDWEASTLTLC-DRDPDP 631
308 LHOEVNLVWMRATOLQKNLTCEVWG---TSPLKMLSLKENKAKVSKREK----- 356
632 VHLQVTVNAKRVVTLN-----WTPGDNNSPILLEYVEFEDQDMKENGBELKRVAD 684
357 -----PVM 359
685 KGHVNLPLM 693


```
Db      216 RAIFYDEAEDNELT-----FKAGEIITVLDDSDPMMWKGETHOGIGLFPSPNFV 264
Qy      229 VEKLTGSELWQARASSSKSMITFDLKNKEVSKRYTQDPK-----LQNGK 276
Db      265 TADLTA-----EPEMITTEKKTVOF---SDVOVETIIEPEEPAFIDEDKMDOLLQW-- 313
Qy      277 KLPLHLT-----LPOALPOYAGSGNLTALAEKGTGLHQ-----EVNLVVMRATQL 322
Db      314 ---LOSTPSSDDQDPLPELHLHLMCHOMGPIDEKLEDIDRKHSLSLNVKWMALSL 370
Qy      323 QKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMGCILS--DSGOVLL 380
Db      371 YTKLMNE-----DPVSMYAKLQN-----QPYVMQSGVSGSYVAGPPSGAVLV 416
Qy      381 ESNIRV--LPTWSTVHPRAS--ALPAPPTGSALPDQTSALPD----- 421
Db      417 AGNQMASHLQSYSLPPEOLSLSOAVVPSANPALPSQOTQAAVNTVWSSVOGNTYPSQ 476
Qy      422 -----PPAASALPALAVISFLGLG 442
Db      477 APVYSPPAATMAATADVTLYQNG 502

RESULT 78
AC3616
succlnnglycan biosynthesis transport protein exop [imported] - Brucella melitensis (stra
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_reviseion 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AC3616
R:DeJvecchio, V.G.; Kaparatl, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivanova,
: Mazur, M.; Goldman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Leese
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AJ03252; PMID:11756688
A:Accession: AC3616
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-729 <KID>
A:Cross-references: GB:AE008918; PIDN:AAU54094.1; PID:q17965053; GSPDB:GN00191
A:Experimental source: strain 164
C:Genetics:
A:Gene: BMEI10852
A:Map position: 11

Query Match      4.9%; Score 113; DB 2; Length 729;
Best Local Similarity 20.9%; Pred. No. 8.8;
Matches 98; Conservative 65; Mismatches 155; Indels 150; Gaps 20;

82 SRSRLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOLVGLTANSPTHLQGSLLT 141
116 SRTVLSQPDRTTLKNAK--DFDPAVRE---RQIDLLSLRLVNSD-----GRSFTI 164
142 TLESPPGSSPVQCRSP-----RGKNIQSGKTLVSQLELDQSGTWTCTVLQNGKVEF 195
165 -----FISYAPDEPVAAVAANAFATAYLVN--HQVDVQGSAAARVSEMLGSKVLT 213
196 KIDI-----VLAFOKASIVYKKEGEVYESFP---LAFVVKLTSGGELMWAERAS 246
214 RNDLETARVADFPKSRLLA---GEOGQISFOAQRVWALTEITVAATGAVSTAQARLQ 269
247 SSK-----SWITFDLKNKEVSVKRVNTQDPKLOMGK---LPLHLTLPOA 287
270 TQALKNNNNEAPAMEILLASPAIONLKNBEARVQHHLDELKANGALKSAEITVLMAERES 329
288 LPQVAGSNLTALAEKGTGLHAEVNLVVMRATOLQKNL----- 326
330 LKQ-----QITAOVDIITKSLSNEIRIAVQRTSLEKELKEAFETDLAKANGAQAQVRAQLD 384
327 -----TCVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGM 367
385 REANASRVVETVLTTRYKQIIEODGIAIPEAQILSQASPMAKASPR-----LIN----- 434
368 WQC-----LLSDSGOVLBSNIVKLPWSTPVPVPRASALPAPPTGSALP--DPQ--TASA 418
```

```
Db      435 WLVLGLGIGLFAVAGTMLREALDKIRPAQT-----ALALPGIPATLPLVPSQLTVPI 489
Qy      419 LPDPPAASALP-----AALAVISFLGLGLGY 445
Db      490 LVNRGVDSASPFGRRAIKSVHDLRLALIRGRDSLALSVSSLEGDGKG 537

RESULT 79
158164
BIG-1 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-Jul-1996 #sequence_reviseion 26-Jul-1996 #text_change 24-Sep-1999
C:Accession: 158164
R:Yoshinara, Y.; Kawasaki, M.; Tani, A.; Tamada, A.; Nagata, S.; Kagamiyama, H.; Mori, K
Neuron 13, 415-426, 1994
A:Title: BIG-1, a new TAG-1/F3-related member of the immunoglobulin superfamily with neu
A:Reference number: 158164; MUID:94338697; PMID:8060619
A:Accession: 158164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1028 <RES>
A:Cross-references: EMBL:U11031; MID:9563132; PIDN:AAA63607.1; PID:9563133
C:Genetics:
A:Gene: Big-1
C:Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homology

Query Match      4.9%; Score 113; DB 2; Length 1028;
Best Local Similarity 19.8%; Pred. No. 14;
Matches 109; Conservative 69; Mismatches 174; Indels 198; Gaps 25;

23 TQGNKVVLGK-----GDVYELTCTASQK--SIGTHMKNSNOKIKLNGSGFLTK-GP 73
497 TEPRIILAPSNDVAVGESIILPCQVQHDPLDIMPAWFFNGTLTFPKKDGSHFEKVG 556
74 SKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEV---QLTVFG----- 124
557 SSSGD-----LMTNQLKSGKVKVCMVQICVDSSAAELIVAGSGPP 601
125 -----LTVNSDTH-----LLOQSLLTLT 142
602 ENVKVDITDTTQALSTEGTDSHSPIYSVAQARTPSVGMQVWRTVPEADIKTRTAT 661
143 -LESPP-----GSSPSVQCRSPRK-----NIOGKTLVSQI--- 174
662 VVELNPWEYEFRRVVASNKIGGEPSPSEKVRTEBAPEVAPSEVGGG--SRSELVIT 720
175 -----ELDQSG-----TWTCTVL---ONQKVEFKIDIVVLAFOKASSIV 211
721 WDPVPELQNGGCGGYVAAPRPLGVTWITQIVTVSPDNPRVFPNBSIVPSPVEVKG 780
212 YKKEGEVYESFPLAFVVEKLTSGGELMWAERASSK--SWITFDLKN----- 258
781 YNNKGEGFPSPVTTVFSAEEPEPTVAPSHISHSLSSEIEVSNMTIWMKSGNRLGYEV 840
259 -----KEYSVK--RTQDPKLOMGKKLPLHLTLPOALPOY--AGSGNLTALAEKGT 307
841 RYMNNGEESBSISVKAQNTSAVLGLKSNLAVYTAVRVNTAGAPSATVNAATTKK 900
308 LHQEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLK---LENKEAKV-----SKRE 355
948 NNVOVLNTNKTSAELLPLIKEDYIIEVKATTDGSDGTSSEQIRI-----PRITSM 998
403 APPTGSALPD 412
999 ARGSTASID 1008

RESULT 80
```



```

Oy 121 LVFGLTANSDPHLQGGSLTLTLESPGSSPSVQCRSPRG-----KN-----IOGKTL 169
Db 133 VVKVKQSSQSEPIHLQADPLETEKLMGECVTRSYEBGNVTYKNGRVLQPVBEVVI 192
Oy 170 SVSOLELQDSG-----TWTCFTVL-----QONKRVF----- 195
Db 193 NLARKVENSTGLFTWTSSLQWPTKEDANNAKFTCIIVTHGPGSQQTIGSEPVFVDVHYPT 252
Oy 196 -KIDVLAFOKASSIVYKKEGEQVEFSPFLAFYTEKLTGSGE-----LWM---QARA 245
Db 253 EKVTRIVLS--QSTT-----KEGDNYTL-----KSGNGNPPQFLEFIPETGEI 298
Oy 246 SSSKSWITFDLKNKEVSVKRVTDPK-----LQMGKCLPLH 281
Db 299 RSSDPTVMTDVRANATGECVKSLLIKSMMDATTIVHLDLQLTSPSGEVTQOIGALVVS 358
Oy 282 LTLPOALP-----QVAGSGL-----TLAEAKTGKLTQEVNLVY- 316
Db 359 CTISSRNATVFWIKDNTRMKTSPSSSLQYODAGNYICETTLQVEGLKKRKTLLIVE 418
Oy 317 -----MRATOLQKLTCEVWGPTSPKMLSLKLENKAKVSKREKPMV 360
Db 419 GKPOIKMTKKTNTNKMSTIVCHVEG--FKPQAVQMTVYSGSLINKTEETKYV 470

```

RESULT 83

glial cell membrane glycoprotein LIG-1 precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Apr-1997 #sequence_revision 11-Apr-1997 #text_change 05-Nov-1999
 C:Accession: A58532

R:Sunuki, Y.; Sato, N.; Tohyama, M.; Manaka, A.; Takagi, T.
 J. Biol. Chem. 271, 22522-22527, 1996

A:Title: cDNA cloning of a novel leucine glycoprotein that is expressed specifically in A:Reference number: A58532; MUID:96394313; PMID:8798419

A:Accession: A58532

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1091 <SUZ>

A:Cross-references: GB:D78572; NID:G1545806; PIDN:BA11416.1; PID:G1545807

C:Superfamily: leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan amino-ter

F:36-61/Domain: proteoglycan amino-terminal homology <PMH>

F:71-94/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>

F:95-117/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>

F:118-141/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>

F:142-165/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>

F:166-189/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>

F:191-213/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>

F:214-237/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>

F:238-261/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>

F:262-285/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>

F:286-309/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR10>

F:310-333/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR11>

F:334-357/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR12>

F:358-381/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR13>

F:385-408/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR14>

F:409-432/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR15>

F:440-485/Domain: proteoglycan carboxyl-terminal homology <PCH>

Query Match 4.8%; Score 112; DB 2; Length 1091;
 Best Local Similarity 21.2%; Pred. No. 18;
 Matches 95; Conservative 59; Mismatches 166; Indels 120; Gaps 21;

```

Oy 14 LQALPLPAATGKNNKVLGKKGDTVELTCTASQKSIQFHWKNSNOIKILNGSFLYKGP 73
Db 595 LTVNVLPSFTKIPIHIAIRGTGTARLECAATGHPNPQIAMQDGG-----TDFP 643
Oy 74 SKLNRADRSRLMDQGNFPLIKLKLESDPTIYCEVEDQKEVQLVFGLTANSDPHL 133
Db 644 A-----ARRRRHNVMPDDVFTTVDKIDMGVYCTQANSASGV-----SANATLTV 691
Oy 134 LOGOSLTLTLES---PPGSSPSVQCR-----SPRKNIGGKTLIS----- 170

```

```

Db 692 LETPSLAFLBEDRVVTVGEVAFQCKATGSPPTPIWLGKGRPLSLTERHHFTPGNQLLV 751
Oy 171 VSOLELQDSGWTCTTVQONQKVEFKIDIVLAQKASSIV-----YKKEQVEFSF 223
Db 752 VQNMIDAGGYTEMEN-----PLGERHQSLSLPTPGCRDGTTVGI-F 798
Oy 224 PLATVEKLTGSGELMMQOERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKCLPLHLT 283
Db 799 TIAVCGSLVLS--LWVVCIIYQTRK-----KSEETSYINDE-----TI 836
Oy 284 LPOALPOYAGS-GNLTALAE--AKTGKLTQ-----EVNLVVRATYLOKRL-----TCV 330
Db 837 VPPDPVPSYLSQGTLSDRQETVVRTEGQHANGHIESNGVCLRPDSLFPVEDIHSTCR- 895
Oy 331 WGPSPKMLSLKLENKAKVSKREKPMVNLNPAAGMOCCLSPSGVLLESNTIKVLP 390
Db 896 ---QPKLCVGYTRE--PWKTERADPTAAPHTTHASGAVSCDC-----ST 936
Oy 391 STPVPRASALPAPPTGALPDPTGASA 418
Db 937 DTAVHPQ---PVRDSG---QPGTASS 957

```

RESULT 84

S46374
 Ig kappa chain V-J region (T33-4) - human (fragment)

C:Species: Homo sapiens (man)
 C:Date: 27-Jan-1995 #sequence_revision 01-Sep-1995 #text_change 21-Jan-2000
 C:Accession: S46374; S38651

R:Bensimon, C.; Chastagner, P.; Zouali, M.
 EMBO J. 13, 2951-2962, 1994

A:Title: Human lupus anti-DNA autoantibodies undergo essentially primary V(chi) gene rea

A:Reference number: S46369; MUID:94313975; PMID:8039491

A:Accession: S46374

A:Molecule type: mRNA

A:Residues: 1-120 <BEN>

A:Cross-references: EMBL:Z27175; NID:G415965; PIDN:CAA81699.1; PID:G415966

C:Superfamily: immunoglobulin V region; immunoglobulin homology

C:Keywords: heterotetramer; immunoglobulin

F:31-105/Domain: immunoglobulin homology <IMM>

Query Match 4.8%; Score 111.5; DB 2; Length 120;
 Best Local Similarity 34.6%; Pred. No. 0.98;
 Matches 37; Conservative 7; Mismatches 52; Indels 11; Gaps 3;

```

Oy 11 LVTVQLALPLPA-----TQGNKVLGKKGDTVELTCTASQKSIQFHWKNSNOIKILGN 64
Db 2 LIGLILLMLPQAKCDVQGTQSPSTLASVSGDRTVITCRAGQSISTWLAWYQKS-----GN 57
Oy 65 QGSFLTGKPSKLNDRADRSRLMDQGNFPLIKLKLESDPTIYCEVEDQKEVQLVFGLTANSDPHL 110
Db 58 APKLLIYGASNLBSGVPSRFSGSGSTFDITLTSSLDPPDSATVYCC 104

```

RESULT 85

I37891
 interleukin-11 receptor alpha chain - human

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 03-Dec-1999

C:Accession: I37891; G01970; G01971

R:Cherel, M.; Sorel, M.; Lebeau, B.; Dubois, S.; Moreau, J.F.; Bataille, R.; Minvielle,

Blood 86, 2534-2540, 1995

A:Title: Molecular cloning of two isoforms of a receptor for the human hematopoietic cy

A:Reference number: I37891; MUID:95399754; PMID:7670098

A:Accession: I37891

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-422 <RES>

A:Cross-references: EMBL:Z38102; NID:G995653; PIDN:CAA86224.1; PID:G995654

R:Van Leuven, F.

submitted to the EMBL Data Library, July 1995

A:Reference number: G08959

A:Accession: G01970

[illegible]

F:101-162/Domain: immunoglobulin homology <IM>
F:408-693/Domain: protein kinase homology <KIN>

Query Match 4.8%; Score 111.5; DB 2; Length 750;
Best Local Similarity 21.6%; Pred. No. 11;
Matches 79; Conservative 41; Mismatches 102; Indels 143; Gaps 16;

QY 15 QALLPATQGNKVKVGGKGDVTELTCTASQKKSIOFHW-KNSNOILKLNQSGFLTKG- 72
DB 94 KLHAYVPA-----NTVRCRCRAGCNPTSPMRMLKNGKEFQEHRIIGGFKYSQ 141
QY 73 -----PSKLNDRADSRSLMDQGNFPLIKKLKIEDSPDYICEVEDQKEVOLLV 122
DB 142 HPSLMIEVSVEB-----DEGNVTCIMENDYGSINHTYHLDYERSHPRPLQ 188
QY 123 FGLTANSPT-----HL-----LQGSLTLTLESPPGSSPSVQCRS 157
DB 189 AGLPANTTTKVGDAEFVCKVYSDAQPHIQWIRHFELNGSKI-----GPDGHYLYATLK 242
QY 158 PRGNKIQQGKTLVSQLELQDSGTWTC-----TYL-ONOKVVEKIDIV 201
DB 243 RSGIINSSNAEVLTLHNVTEDAGQYTCVKVSNYIGANQSAMLTYLPAPEKDEBERELD-- 299
QY 202 LAFQKASSIV-----KKEGQVFEFSPLAFTVTKLNGSG 236
DB 300 -----SSSEYELIAYCYGGLITCMITIMVCHMKGRKQSDFFSPPA-VHKLISKL 351
QY 237 ELWMQ-----AERASSKS-----WTFDLKNKE-----VSVKRVYQDP-----KLQ 273
DB 352 PLRQVTVASADSSSMNSMTPLVRIITRLSSNNDDTHLGVSVSEYELPEDEKMEYPREKLT 411
QY 274 MGKKL 278
DB 412 LGKPL 416

RESULT 87

S03517
T-cell receptor gamma chain precursor (clone PM17664) - human
C:Species: Homo sapiens (man)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 21-Jan-2000
C:Accession: S03517
R:Tighe, L.; Forster, A.; Clark, D.M.; Boylston, A.W.; Lavenir, I.; Rabbits, T.H.
Eur. J. Immunol. 17, 1729-1736, 1987
A:Title: Unusual forms of T cell gamma mRNA in a human T cell leukemia cell line: implic
A:Reference number: S03517; MUID:88083067; PMID:2961573
A:Accession: S03517
A:Molecule type: mRNA
A:Residues: 1-340 <TIG>

A:Cross-references: EMBL:X06774; NID:g37346; PIDN:CAA29941.1; PID:g37347
A:Note: This sequence was determined from the differentiated gene
C:Superfamily: Immunoglobulin V region; immunoglobulin homology
C:Keywords: T-cell receptor
F:1-30/Domain: signal sequence #status predicted <SIG>
F:21-340/Product: T-cell receptor gamma chain #status predicted <MAT>
F:21-116/Domain: V region (V-gamma-8) #status predicted <VR>
F:121-135/Domain: J region (J-gamma-2) #status predicted <JR>
F:136-340/Domain: C region (C-gamma-2) #status predicted <CR>
F:160-225/Domain: immunoglobulin homology <IM>

Query Match 4.8%; Score 111; DB 2; Length 340;
Best Local Similarity 20.9%; Pred. No. 4.3;
Matches 77; Conservative 58; Mismatches 138; Indels 96; Gaps 16;

DB 10 LLIVQLALLPAAQGN-----KVVLGKKDVTVELTCTASQKKSIOFHMKNSNOIKIL 62
DB 3 LALALLAFLPPASQKSSNLBGRKTSVTRPTGSSAVITCDLPVENAYTTW-----YL 55
QY 63 GNQGSFLLTKGPKSL--NDRADSRSLMDQ-----NPLIKLKLIEDSDT 106
DB 56 HQEG-----KAPQRLIYDYSNRYVL-ESGISREKHYTAATGSKLFILENLIERBSGV 110
QY 107 YICEVEDQKEVOLLVFGLTANSPTHLLOGQSITLT--LESPPGSSPSVQCRSPRGKNI 163

```

Db      111 YCATWTDRYYKKL-FG-----SGTLVVTDRQDLADADSPKPTIFLP----- 152
Qy      164 QGCKLSSQLELDQSGTCTVLOKQKVEFKIDIVLAFQ--KASSIVYKKEGE----- 217
Db      153 -----SIAETLQKAGTYLCLL-----EKFFPDIKHKMOEKSMITLGSQEBNTMKT 200
Qy      218 ---QVEFSPLAFYTEKLTGSGELMWQAEARASSKSMITFDKKNKEVSKVATODPKLQM 274
Db      201 NDTYKFSW-LTVPEBSLDKEHRC--LYRHENKNGIDQEIIFPKITDVTTPKOSY 256
Qy      275 GK-----KLPRLHLLTPQALPQYAGSGNLTALAEAKTGKLGHOEVLNVKRA 319
Db      257 SKDANDVTVDKPNVSKDANDVTMTDPKDNWSKDANDTLLQLNTSAVYVYLLLLKS 316
Qy      320 TOLQKLTG 328
Db      317 VVYFAITTC 325

```

RESULT 88

```

cell adhesion molecule gp80 - slime mold (Dictyostelium discoideum)
A44100
C:Species: Dictyostelium discoideum
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
C:Accession: A44100
R:Debarats, L.; Lam, T.Y.; Wong, L.M.; Siu, C.H.
J. Biol. Chem. 267, 19655-19664, 1992
A>Title: Identification of a unique CAMP-response element in the gene encoding the cell
A:Reference number: A44100; MUID:92406924; PMID:1326559
A:Contents: AX2
A:Accession: A44100
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-514 <DES>
A>Note: sequence inconsistent with the nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBIP:113993)

```

Query Match 4.8%; Score 111; DB 2; Length 514;

Best Local Similarity 22.0%; Pred. No. 7.5; Matches 96; Conservative 60; Mismatches 181; Indels 100; Gaps 22;

```

Qy      63 GNGSFLTKGPKLNDRAISRSLWDQGNFPLIKLKIEDS-DTYICEVEQKEEYQL 121
Db      116 GRIGMLNDGPNISG-----YKLNVDNSINAMLSV--TADSVSPT 155
Qy      122 VEGLTANSPTLLQGSLLTLESPPGSSPVQCRS-----PRGKNIQ 164
Db      156 IYPLVNT-----IAGGLNLLELIQFPGFSTIVTSKVSFSPITISITPLAFDLTPNVT 211
Qy      165 GG---KILSVSQ-----LELDQSGTCTVLOKQKVEFKIDIVLAFQKASSIYK 213
Db      212 GRYFVTTASVTWGSHTYGLTVQDDGT-NCHVIFTRSYESSNITITAKASTGVDMITLD 270
Qy      214 KEQGEVFEFSF---PLAFTVEKLTGSGELMWQAEARASSKSMITFDKKNKEVSKVATOD 269
Db      271 NQGNQOPITFTYNPPTITSTKQVNDSEI-----STNTGTDF---TQSLTMGTSS 319
Qy      270 PK-LQMGKRLPLHLLTPQALPQYAGSGNLTALAEAKTGKLGHOE-VNLVVMRATOLQK 324
Db      320 PNLVITGTNEKIVITLPHALPE---GRIGFLNKAISNVVTSITLVTPVINSTVQAPH 375
Qy      325 N-LTEWVGPTSPKMLSLKLENKEAKV-----SKREKVVWLNDEAGCMQCLSDSQV 378
Db      376 NQGSITISGIFLNNAHVSIVDQNTTDIYCAPDSNGESITCFVDAGSGINILVLTNYKNF 435
Qy      379 LLESNIKVLPTWSTPV-----HPRASALPAP-PTGSALPDP-----OTASALPDPAPAS 426
Db      436 ASDPITKTKATSTTYITDTPPTDTPATPSPTPTDTPATPSPTPTSTPETEAPSSAT 495
Qy      427 AL-PAALAV--ISPLL 439
Db      496 TLISPLSLIVIFISFVL 512

```

RESULT 89

```

A45254
surface glycoprotein BEN precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 21-Jul-2000
C:Accession: A45254; #19202
R:Pourquie, O.; Corbel, C.; Le Caer, J.P.; Rosier, J.; Le Douarin, N.M.
Proc. Natl. Acad. Sci. U.S.A. 89, 5261-5265, 1992
A>Title: BEN, a surface glycoprotein of the immunoglobulin superfamily, is expressed in
A:Reference number: A45254; MUID:92302224; PMID:1608932
A:Accession: A45254
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-588 <POU>
A:Cross-references: EMBL:X64301; NID:963087; PIDN:CAA45579.1; PID:963088
C:Keywords: glycoprotein

```

Query Match 4.8%; Score 111; DB 2; Length 588;

Best Local Similarity 19.2%; Pred. No. 8.9; Matches 91; Conservative 68; Mismatches 175; Indels 140; Gaps 19;

```

Qy      8 RHLVLVLQALPLPATQGNKRVLGKGPVELTCTASQKSIQF-HMK---NSNQIKIL 62
Db      16 RPLCLLLAALCWPRLAGLYTVNAVYGDITMPCRLLEVPDGLMGKMYEMPNSPVPIA 75
Qy      63 GNGSFLTKGPKLNDRAISRSLWDQGNFPLIKLKIEDSDTYICEV--EDQKEEYQL 120
Db      76 FRSS---TKKNVQYDDVPDYDRSLSENLYTSLIKNARISDEKRFVCMLVTEDDVSEPT 132
Qy      121 LVFGLTANSPTLLQGSLLTLESPPGSSPVQCRSPG-----KN-----IQGKTL 169
Db      133 VVKVFKQSPQPEILHQADFLTEKLMKLGCVVDSIPBGNVTYKGRVLIQVBEVVYI 192
Qy      170 SVSQLELDQSG-----TWCTVL---QNKQVEF----- 195
Db      193 NLRKVENRSTGLFMTSSLOQWPFKEKANAKFTICIVTHGSGQKTIQSEVVDVNHPT 252
Qy      196 -KIDIVLAFQKASSIYKKEGEVFSPLAFTVEKLTGSGE-----LW---QAEKA 245
Db      253 EKVITRVLS--QSSFI---KEGDVNTL-----KCSGNGNPPQFFLYIPGETEGE 298
Qy      246 SSSKSMITFDKKNKEVSKVATODPK-----LQMGKRLPLH 281
Db      299 RSSDTYVNTDVRNATGEYKCSLIDKSMDDTTTTHVLDQLTPSGEVTQIGBALPVS 358
Qy      282 LTLPOALP-----QYAGSGNL---TLAEAKTGKLGHOEVLNV- 316
Db      359 CTISSRNATVFWIKDNTRMKTSFSSLOQYQDAGNYICETTHKEVEGLKGRKTKLIVE 418
Qy      317 -----MRATOLQKNLTCCEVWGPTSPKMLSLKLENKEAVSKREKPVWV 360
Db      419 GKQIKMTKKTNTNMKSTIVCHVEG--FPKPAQVMTVTSGLSINKEETKYV 470

```

RESULT 90

```

C81412
NOL1/NO2/sun family protein Cj0636 [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: C81412
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Bauman, D.; Chillin,
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A>Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: C81412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-279 <PAR>
A:Cross-references: GB:AL139075; GB:AL111168; NID:96967817; PIDN:CAB75272.1; PID:9696810
A:Experimental source: serotype O2, strain NCTC 11168

```



```

QY      23  TQGNKVVJGK-----GDTVEITCTASQK--SIOFHKNSNOIKILGNQSPFLTK-GP 73
      497  TEPTRIILAPSNMVAIVESVILPCOVQDPLDIDFAMYPFGALTDPFKQDSHREKYG 556
QY      74  SKLNDRADSRRLMDQGNPLIIKNLIKEDSDTYICEVEDQKEV-----QLLVFG----- 124
      557  SSSGD-----LMIRNQLKHSKGYCMQGTGDSVSSAAELLVRGSPGP 601
QY      125  -----LTRANSDH-----LLQGSLTIT 142
      602  ENVKVDEITDTTAQLSWTEGTDHSHSPVISYAVQARTPFSGVQSVRTVEVIDGKTHAT 661
QY      143  -LESP-----GSSPVGCRSPRGK-----NIGGKTLVSGQL----- 174
      662  VELLPWVEYEFRIYASNKIGGGEPSLSEKVRTEAEIAPSEVSGGG-SRSELYTT 720
QY      175  -----ELQDSG-----TMTCTVL-----QNKVVEFKIDIVLAFORASSIV 211
      721  WDPVEELQNGGFGYVAFRPLGVTWTIQTIVTSPDNRVYFRNESIVPSPRYEKYGV 780
QY      212  YKKEGQVFFSPPLAFTVEKLTGSGELMQAERASSK--SWITFDLK----- 257
      781  YNNKGBGPPSPYTTVFSABEEPTVAPSHISASLSSEIEVSMNTIPWKLNGHLGVEV 840
QY      258  -----NKEVSVKRV-----TODPKLQMGKKLPLHLTLPOALPOY--AGSGNLTLLAKT 305
      841  RYNNNGGEESSRRKVKVAGNQTSAVLRG--LKSNNLAVYTAVPAYNSAGPPSATVNATT 898
QY      306  GKLEHGVNLVWVRATQQLKLTCEVWGPTSPKLMLS--LKLENKAKV-----SKR 354
      899  KK-----TPSQPPENV--VMNATDTKVLNMEQVXAMENSEVTGKVFYRTSS 946
QY      355  EKPWVVLNPEAGMOCCL-----SDSGVLLIESNIKVLPTWSTVHPRASAL 401
      947  QNNVHLNNTKTSABELLPIKEDYIIEVKATTDGDTGTSSEQIRI-----PRISM 997
DB      402  PAPPTGSAL 410
QY      998  DARGSTISAI 1006

```

RESULT 94
A41228
protein-tyrosine kinase (EC 2.7.1.112) Flk-1 precursor, endothelial cell-specific receptor
C:Species: Mus musculus (house mouse)
C>Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 04-Feb-2000
C:Accession: A41228; A46065; I58365; S18832; S29991
R:Matthews, W.; Jordan, C.T.; Gavin, M.; Jenkins, N.A.; Copeland, N.G.; Lemischka, I.R.
Proc. Natl. Acad. Sci. U.S.A. 88, 9026-9030, 1991
A>Title: A receptor tyrosine kinase cDNA isolated from a population of enriched primitive
A:Reference number: A41228; MUID:92020984; PMID:1717995
A:Accession: A41228
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1167 <MAT>
A:Cross-references: GB:X59397; NID:950976; PIDN:CAA2040.1; PID:950977
R:Willauer, B.; Wilmann-Voos, S.; Schunrich, H.; Martinez, R.; Moller, N.P.; Rissau, W.;
Cell 72, 835-846, 1993
A>Title: High affinity VEGF binding and developmental expression suggest Flk-1 as a major
A:Reference number: A46065; MUID:9308880; PMID:7681362
A:Accession: A46065
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-24, 'T', 26-782, 'VL', 785-916, 'C', 918-1367 <ML>
A:Cross-references: GB:X70842; NID:957923; PIDN:CAA50192.1; PID:957924
A>Note: submitted to the EMBL Data Library, January 1993
A>Note: sequence extracted from NCBI backbone (NCBI:128064)
R:Oelrichs, R.B.; Reid, H.H.; Bernard, O.; Ziemielski, A.; Wilke, A.F.
Oncogene 8, 11-18, 1993
A>Title: NYK/Flk-1: a putative receptor protein tyrosine kinase isolated from E10 embryo
A:Reference number: I58365; MUID:93141255; PMID:8423988
A:Accession: I58365

```

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-678, 'D', 680-1340, 'RSPPV', <DEL>
A:Cross-references: GB:S53103; NID:9264004; PIDN:AAE25043.1; PID:9264005
C:Genetics:
A:Gene: FLK-1; NYK
C:Superfamily: unaassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo-
C:Keywords: ATP; autophosphorylation; phosphoprotein; phosphotransferase; transmembrane
F:830-1165/domain: protein kinase homology <IN>
F:838-846/Region: protein kinase ATP-binding motif

```

Query Match 4.7%; Score 109; DB 2; Length 1367;
Best Local Similarity 21.0%; Pred. No. 38;
Matches 45; Conservative 36; Mismatches 73; Indels 60; Gaps 9;

```

QY      9  HLLVLQALLPATQGNKVVLGKKGDTVEITCTASQKSIQFHKNSNOIKILGNQSGF 68
      544  HVIRGPEITVQPAQPTQ-----ESVSLCTADNRFTFENLTWK-----LGSQATS 590
QY      69  LTKGSPKLNDRADSRSLW-----DQGNPLII--KNLIKEDSDTYICEVEDQKEE- 117
      591  VMGBS-LTPVCNKLDALMKLNGTWFSNTDILVAFQNASLQDQGYVCSAQDKTKK 649
QY      118  -----VQLLVGGLTRANSPTHLLQGSLTTLTLESPPGSSPYQCRSPRGK----- 162
      650  RHCLVKQIILIERAPMTITGNLENQTTI-----GETIEVTC--PASGNPPIHTWFKD 701
DB      163  -----IQGKTLVSVQLELQDSGTWTC 184
      702  NETLVEDGIVLRDGNRLITRRVAKEDGGLYTC 735

```

RESULT 95
I50478
neurolin - goldfish (fragment)
C:Species: Carassius auratus (goldfish)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996
C:Accession: I50478
R:laessing, U.; Giordano, S.; Stecher, B.; Lottepeich, F.; Stuermer, C.A.
Differentiation 56, 21-29, 1994
A>Title: Molecular characterization of fish neurolin: a growth-associated cell surface p
protein DM-GRASP/SC-1/BEN.
A:Reference number: I50478; MUID:94299040; PMID:8026643
A:Accession: I50478
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-523 <LAB>
A:Cross-references: GB:L25056; NID:9407318; PID:9407319

Query Match 4.6%; Score 108; DB 2; Length 523;
Best Local Similarity 18.7%; Pred. No. 12;
Matches 73; Conservative 68; Mismatches 127; Indels 122; Gaps 20;

```

QY      34  GDTVEITCTASQKS--IOFHKNSNOIKILGNQSPFLTKGPK-----LNDRADSR 84
      2  GETIVPNDGTGKRPDGLIFTRK--YKDDSGEGLLVQAQKDEATVATDGYKSRV 58
QY      85  SLMDQGNPLIIKNLIKEDSDTYICEVEDQKEEVQLVFGLTANSPTHLLQGSLTTLTLE 144
      59  SI--AANSLLIARGSLADQRFVC-----MVVSFTN-----LREYSVEYVH 99
QY      145  SPGSSPEVQCRSPRGKNIQGGKTLVSVQLELQDSGTWTCVTLVNQKNVEKIDIVLAF 204
      100  KKP-SAPYIK--NNAKLENGKLTQLECEVENA----- 130
QY      205  QKASSIVYKKEGQV-----EFSPLAFTVEKLTGSGELMQAERASSKSWITFDLK 259
      131  NPPADLIYKKNQTLVDDGKIIITITTKTKTG-----LSTSSRLQYTAKE 180
QY      260  EVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLLAKTKGLH--QEVNLIV 316
      181  DVE-SQFCTAKHWG--PDVSEPSFP-----IHYPTEKVISQV 218

```

Qy 317 MRATOLQK-----NLTCVWGPSTPKMLSLKLENKAVSKREKPVWL-----NPEAGMW 368
 Db 219 VSQSPTRRGEVDYTLKQADGNPP---TSFNNIKGKKVYTKDQVYTLTGVTAADSGIT 275
 Qy 369 QCLSDSGQVLES-----NIKVLPT 389
 Db 276 KCSLNDND--VWESQFVTSFLDVSFLTPT 303

RESULT 96

I68093
 C:Species: Homo sapiens (man)
 C:Accession: S40370
 C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jan-2000
 R:Klein, R.; Jaenichen, R.; Zachau, H.G.
 E:Eur. J. Immunol. 23, 3248-3271, 1993
 A:Title: Expressed human immunoglobulin chi genes and their hypermutation.
 A:Reference number: S40312; MUID:94080891; PMID:8258341
 A:Accession: S40370
 A:Status: preliminary; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-122 <KLE>
 A:Cross-references: EMBL:X72480; NID:9441428; PID:CAA51148.1; PID:9441429
 C:Superfamily: immunoglobulin V region; immunoglobulin homology
 C:Keywords: heterotrimer; immunoglobulin
 F:30-104/Domain: immunoglobulin homology <IMM>

Query Match 4.6%; Score 108; DB 2; Length 538;
 Best Local Similarity 21.3%; Pred. No. 13;
 Matches 112; Conservative 58; Mismatches 204; Indels 152; Gaps 25;

Qy 2 NRGVFRLLLVQLALPAATQGNK-----VLGKGDVTELTCTASOKSIQPHMKNNOIKILGN 41
 Db 10 SRSPETPLMLPLLLLETTGAQDVAVYLBVRGQLGTVELPCHLLPVPGLYISLVT 69
 Qy 42 -----TASOKKSIQPHMKNNOI---KILGNQSGFLTKGSKLNDRADSRSLMDQGNF 92
 Db 70 WQRPAPANHONVAFHFKMGSPSPKPSERLSFVS---AKOSTGDTAEALODA--- 123
 Qy 93 PLIINKLIEDSDTYICEVD--QKEEVO---LIVFGLTANSSTHLLQGQSLTTLTLESPP 147
 Db 124 TLALHGLTVEDEGNTCEFAFEPKGSVGRMTWLRIYAKPKN-----QAEAKVTFSQDP 177
 Qy 148 GSSPEVQGRSPRGK---NIQSGKTLVSQLELDQSGMTCTVYLQNKVEKIDIVLAF 204
 Db 178 --TIVVATLISKEGRPARISWLSLIDWEAKETQVSGTLAGTV---TVTSRFTLVPSGR 230
 Qy 205 QKASSIVYKKEGEQVE--FSFPLAFTVE-----KLTSGSELWMQAE-----RASSS 248
 Db 231 ADVGVTVCKVEHSEFEPEPALIPVLTLSVYRPREVSISSGIDDMWYLGRTATATLSCVRSNP 290
 Qy 249 KSWITFDLKNKEVSVKRYTDPKLOMGKPLH-----LTLPOAL----- 288
 Db 291 PTGYDMSTTSGTFPSAQAQ-----GSQLVIAHADSLENTTFVCTVNAVAGMGRAEQVI 344
 Qy 289 -----FOYAGSGNLTALAEAKTGKHQEVNLVWKAATOLQKNLTC----- 328
 Db 345 FVRETPNTAGAG-----ATGGIIGIITAIATVAVATGILLCRQQRKEQTLQGAED 397
 Qy 329 -EVMGPTSPKMLSLKLENKAVSKREKPVVNLPEAGMQCLLS---DSGQVYLEENI 384
 Db 398 EDLBERP-----SYKPPTPKAKLEAQEMPSQLFTLGSEHSPLKTPYFDAGASTGEEM 451
 Qy 385 ---KVLPTW---STVHPAPASALPAPPTGSALPDFQTASALPDP 424
 Db 452 PRYHELPTLEERSGPIHGAATSL-----GSPFVY-----RGPPA 486

RESULT 97
 S40370
 Ig kappa chain - human

C:Species: Homo sapiens (man)
 C:Date: 06-Mar-1994 #sequence_revision 26-May-1995 #text_change 21-Jan-2000
 C:Accession: S40370
 R:Klein, R.; Jaenichen, R.; Zachau, H.G.
 E:Eur. J. Immunol. 23, 3248-3271, 1993
 A:Title: Expressed human immunoglobulin chi genes and their hypermutation.
 A:Reference number: S40312; MUID:94080891; PMID:8258341
 A:Accession: S40370
 A:Status: preliminary; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-122 <KLE>
 A:Cross-references: EMBL:X72480; NID:9441428; PID:CAA51148.1; PID:9441429
 C:Superfamily: immunoglobulin V region; immunoglobulin homology
 C:Keywords: heterotrimer; immunoglobulin
 F:30-104/Domain: immunoglobulin homology <IMM>

Query Match 4.6%; Score 107.5; DB 2; Length 122;
 Best Local Similarity 35.5%; Pred. No. 1.6;
 Matches 38; Conservative 5; Mismatches 53; Indels 11; Gaps 3;

Qy 11 LVIVQLALPA-----TQGNKVLGKGDVTELTCTASOKSIQPHMKNNOIKILGN 64
 Db 1 LIGLLMLKAGACDQLTQSPSSLSASVGDVRYTTCRASQSTFLHWQON---LGK 56
 Qy 65 QGSFLTKGPKSLNDRADSRSLMDQ--NFPILIKNLKIEDSDTYICE 110
 Db 57 APKLIRYAAANLQSGVPSRPSGSGSDPFLTITSGIQPEDFATYYCQ 103

RESULT 98

S68177
 C-CAM2a protein isoform precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
 C:Accession: S68177
 R:Lucka, I.; Cichocka, I.; Baemmler, K.; Bechler, K.; Reutter, W.
 E:Eur. J. Biochem. 234, 527-535, 1995
 A:Title: A short isoform of carcinoembryonic antigen-related rat liver cell-cell adhesion
 A:Reference number: S68177; MUID:96128184; PMID:8536699
 A:Accession: S68177
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-458 <LUC>
 A:Cross-references: EMBL:X91137; NID:91160272; PIDN:CAA62577.1; PID:91160273
 F:1-138/Domain: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-terminal
 F:1-33/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEA>
 F:1-33/Domain: signal sequence #status predicted <SIG>
 F:34-458/Product: C-CAM2a protein isoform #status predicted <MAT>
 F:252-301/Domain: immunoglobulin homology <IMM>

Query Match 4.6%; Score 107.5; DB 2; Length 458;
 Best Local Similarity 19.2%; Pred. No. 11;
 Matches 101; Conservative 80; Mismatches 175; Indels 169; Gaps 28;

Qy 3 RG-VFPRHLLLVQL-----ALPAATQGNKVLGKGDVTELTCTASOK 46
 Db 10 RQGIPLWRGLLTLTSLTYWSPLTQAQTVDAVPPNVVEKSVLL-----LAHNLPOE 61
 Qy 47 KSTQFHMKNSNOIKILGNQSGFLTK-----GPKSLNDRADSRSLMDQGNPPLIYKILKI 101
 Db 62 FOV-FYVYKGTTL-----NPDSERARYRSDNMSKTGPVAGRETIYNSG--LFFQVYVK 114
 Qy 102 EDSPTYICEVDO-----KEVQLVVE-----GLTANSSTHLLQGQSLTTLTLESPPGSS 150
 Db 115 TDERAYTLVDFDOQFNFIQTSVQFRVYPAIQKNVNGN-NSNPVEGEPP-VSLACEPYTN 172
 Qy 151 PSYQCRSPRGKNI-----QGKTLVSQLELDQSGMTCTVQLN---QKVEEKID 198
 Db 173 NTSYLMWRNGBESLSEGRVTFPSGNRTLTLNVRRTDKGYEECAARNPATFNREDPFLND 232
 Qy 199 IV-----VLAFOKASSIVYKKEGEQVEFSF-----PLA---FTYEKL-TGSGELWMQ 242
 Db 233 VIVGPDAFVI-----SPDILVHQSGNINLSCHADSNNPAGYFWLINELKQTLTSSQELFISN 288

```

OY 243 BRASSKSMITFDLKN-----KEVSVKRYTQDPKLGKGLPLHLTPQALPYA 292
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 289 ITTNNSGYACF-VNNVTYGLSRTTVKNITFEFVTPQSIQITNTTVKEL----- 337
OY 293 GSGNLTALAEK-TGKLHGEVNLVYMRATOLQKMLTCEWGPSPKML-----SLKLENK 347
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 338 --GSVTLTLCFSKDTG-----VSVR-----MLFNSQSLQTLDRMTLSQDNS 375
OY 348 EAKVS--KREKPVWVNLNPEAGMOCCLSDSGQVLLSINIKLPTWSTVPHRASALPAP 405
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 376 TLRIDPIKRE-----DAGDYQCEISN-----P 397
OY 406 TGSALPDQOTASALPDPPASALPALAVISFLGLGIGVACVLA 450
Db 398 VSRFISHPIKLDVTPDPTQNSGLSEGAIVIGSVAGVALIAA 442

RESULT 99
A44783
ecto-ATPase precursor - rat
N.Alternate names: canalicular bile acid transport protein
C.Species: Rattus norvegicus (Norway rat)
C.Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 23-Jul-1999
C.Accession: A44783; A44410
R.Lin. S.H.; Guidotti, G.
J. Biol. Chem. 264, 14408-14414, 1989
A.Title: Cloning and expression of a cDNA coding for a rat liver plasma membrane ecto-AT
A.Reference number: A44783; MUID:89340561; PMID:2527235
A.Accession: A44783
A.Molecule type: mRNA
A.Statut: preliminary
A.Residues: 1-519 <LIN>
A.Cross-references: GB:J04963; NID:g203989; PIDN:AAA41104.1; PID:g203990
R.Stippl, C.J.; Suchy, F.J.; Ananthanarayana, M.; Perlmutter, D.H.
J. Biol. Chem. 268, 2083-2091, 1993
A.Title: The rat liver ecto-ATPase is also a canalicular bile acid transport protein.
A.Reference number: A44410; MUID:93313966; PMID:8420979
A.Accession: A44410
A.Molecule type: protein
A.Residues: 110-120;122-138;148-150 <SIP>
C.Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termi
C.Keywords: glycoprotein; liver; phosphoprotein; transmembrane protein
F.1-138/Domain: carcinoembryonic antigen precursor amino-terminal homology <CEAN>
F.150-217/Domain: immunoglobulin homology <IMM1>
F.252-301/Domain: immunoglobulin homology <IMM2>
F.337-394/Domain: immunoglobulin homology <IMM3>

Query Match 4.6%; Score 107.5; DB 2; Length 519;
Best local similarity 19.2%; Pred. No. 13; Indels 169; Gaps 28;
Matches 101; Conservative 80; Mismatches 175;

OY 3 RG-VFRLHLVLVLOI-----ALLPATQGNKVLGKGDVTELTCTASQK 46
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 10 RQGIQWRGILLTASLLTYWSPLTQAQVTVDAVPRVVEEKSYLL-----LAHNLQOE 61
OY 47 KSIQHWKNSQIKLKGQSFLLTR-----GSKLANDRADSRSLMDQGNPPLIKLKI 101
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 62 FOV-FYWKGTTL-----NPDEIARIYRSNMSKGPAYSGEETIYSGS--LFQNNVK 114
OY 102 ESDSYICEVEDQ-----KEEYQLLVF-----GLTANSDFHLQGSULTLLESPSS 150
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 115 TDERAYTISVFDQGNPIQTSIVQFRVYPAQCKPNTGN--NSNPMGEPP--VSLMCEPFTN 172
OY 151 PSVQGRSPRGKNI-----QGKTLVSQLELDQSGTWTCTVLQN---QKVEFKID 198
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 173 NTSYLMNRNGESLSGDRVTFSEGNRTLLNVRTDGGYECARNPATFNRSPFLD 232
OY 199 IV-----VLAFOKASSIVYKKEGQVEPSF-----PLA-----FTVEKL-TGSGGLMQQA 242
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 233 VIYGDAPVI-----SPPIYLIHQSNLNLSCHADSNPAPQAYFWLINELKQTSQSLFISN 288
OY 243 ERASSKSMITFDLKN-----KEVSVKRYTQDPKLGKGLPLHLTPQALPYA 292

```

```

Db 289 ITTNNSGYACF-VNNVTYGLSRTTVKNITFEFVTPQSIQITNTTVKEL----- 337
OY 293 GSGNLTALAEK-TGKLHGEVNLVYMRATOLQKMLTCEWGPSPKML-----SLKLENK 347
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 338 --GSVTLTLCFSKDTG-----VSVR-----MLFNSQSLQTLDRMTLSQDNS 375
OY 348 EAKVS--KREKPVWVNLNPEAGMOCCLSDSGQVLLSINIKLPTWSTVPHRASALPAP 405
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 376 TLRIDPIKRE-----DAGDYQCEISN-----P 397
OY 406 TGSALPDQOTASALPDPPASALPALAVISFLGLGIGVACVLA 450
Db 398 VSRFISHPIKLDVTPDPTQNSGLSEGAIVIGSVAGVALIAA 442

RESULT 100
S24066
protein-tyrosine kinase (EC 2.7.1.112), receptor type tle precursor - human
C.Species: Homo sapiens (man)
C.Date: 19-Feb-1994 #sequence_revision 11-Aug-1995 #text_change 21-Jul-2000
C.Accession: S24066; C38269; I52613
R.Partanen, J.; Maekelae, T.P.; Alitalo, R.; Korhonen, J.; Sandberg, M.; Renkonen, R.;
Mol. Cell. Biol. 12, 1698-1707, 1992
A.Title: A novel endothelial cell surface receptor tyrosine kinase with extracellular ep
A.Reference number: S24066; MUID:92195316; PMID:1312667
A.Accession: S24066
A.Molecule type: mRNA
A.Residues: 1-1138 <PAR>
A.Cross-references: EMBL:X60957
R.Partanen, J.; Maekelae, T.P.; Alitalo, R.; Lehtvaeslahti, H.; Alitalo, K.
Proc. Natl. Acad. Sci. U.S.A. 87, 8913-8917, 1990
A.Title: Putative tyrosine kinases expressed in K-562 human leukemia cells.
A.Reference number: A38268; MUID:91062389; PMID:2247464
A.Accession: C38269
A.Statut: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A.Molecule type: mRNA
A.Residues: 981-1034 <PMW>
A.Experimental source: clone JTK14
R.Korhonen, J.; Lahtinen, I.; Halmekyto, M.; Alhonen, L.; Janne, J.; Dumont, D.; Alitalo
Blood 86, 1828-1835, 1995
A.Title: Endothelial-specific gene expression directed by the tie gene promoter in vivo.
A.Reference number: I52613; MUID:95383653; PMID:7655012
A.Accession: I52613
A.Statut: translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-19 <RES>
A.Cross-references: GB:S79347; NID:g1086921; PIDN:AAD14299.1; PID:g4261999
C.Genetics:
A.Gene: GDB:TTF, JTK14
A.Cross-references: GDB:212873; OMIM:600222
A.Map position: 1p34-1p33
C.Function:
A.Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C.Superfamily: protein-tyrosine kinase, receptor type tie; EGF homology; fibronectin typ
C.Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F.1-121/Domain: signal sequence #statut predicted <SIG>
F.122-118/Domain: product: protein-tyrosine kinase, receptor-type tie #statut predicted <MAT>
F.215-255/Domain: immunoglobulin homology <IM1>
F.259-302/Domain: EGF homology <EG1>
F.306-344/Domain: EGF homology <EG3>
F.365-428/Domain: immunoglobulin homology <IM2>
F.449-530/Domain: fibronectin type III repeat homology <FN3A>
F.542-633/Domain: fibronectin type III repeat homology <FN3B>
F.642-730/Domain: fibronectin type III repeat homology <FN3C>
F.761-786/Domain: transmembrane #statut predicted <TMW>
F.837-1114/Domain: protein kinase homology <KIN>
F.843-853/Region: protein kinase ATP-binding motif
F.843-105, 372-426/disulfide bonds: #statut predicted
F.83, 167, 503, 596, 709/Binding site: carbonyl/Asn (covalent) #statut predicted
F.870, 887, 979/Active site: Lys, Glu, Asp #statut predicted

```

Query Match 4.6%; Score 107.5; DB 1; Length 1138;
Best Local Similarity 20.9%; Pred. No. 37;
Matches 83; Conservative 44; Mismatches 136; Indels 135; Gaps 18;

```
QY 109 CEVEDQKEEVQLVVGGLTANSDTHLQGSILTLIESPP-----GSSP-----SVQCRS 157
      |||:::|
      |||:::|
Db 344 CEKSDRIPOI-----LNMASELEFNLETWPRINCAAGNPPVPRGSIELRK 389
      |||:::|
      |||:::|
QY 158 PRGNIIQGGKTL-----SVSOLEIDPSGTWCTVLON--QKVEFKIDIVLAFQ 205
      |||:::|
      |||:::|
Db 390 PDGTVLSTKAIVEEKTAEFEVPRVLADSGFEWECRVSTSGGDSRRFKVNVKVPVP 449
      |||:::|
      |||:::|
QY 206 KASSIVYKKEGQVEFSFPLAFVVEKLTGSGELMWQAEERASSSKSWIT-----F 254
      |||:::|
      |||:::|
Db 450 LAAPRLITKQSRQLVSPVFSFGDPISIVRLHYRPQ--DSTMDWSTIVVDPSENVTIM 507
      |||:::|
      |||:::|
QY 255 DLXNK-----EVSVKR-----VTQD---PKLQW-----GKKLPLHLT 283
      |||:::|
      |||:::|
Db 508 NLRPKTGYSVRVQUSRPGEGGEGANGPPTLMTTDCPEPLQPWLEGMHVEGTDRLRVSW 567
      |||:::|
      |||:::|
QY 284 LPQALPQYAGSGNLTLEAKTGKLGHEVNLVVMRATQLOKNIPT-----CEV 330
      |||:::|
      |||:::|
Db 568 LPVVGPIVGDGFLRLMDGTGRGDERRENVSPOARTALLTGITGTHYQLDVQLYHCTL 627
      |||:::|
      |||:::|
QY 331 WQPTSPKLMLSLKENKEAKVSKREKPYWVILNPERG-----MMQCLISDSGQVLLBSNI 384
      |||:::|
      |||:::|
Db 628 LGPASP-----PAHVLLPPSGPPAPRRHLHAQALSDS----- 658
      |||:::|
      |||:::|
QY 385 KVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDP 422
      |||:::|
      |||:::|
Db 659 EIQLTWK--HP--EALPG-PISKYVEVQVAGAGADP 690
      |||:::|
      |||:::|
```

Search completed: August 3, 2004, 13:15:03
Job time : 18.4109 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:50:44 ; Search time 6.2213 Seconds
(without alignments)
3791.557 Million cell updates/sec

Title: SEQ5
Perfect score: 2325
Sequence: 1 MNRGVPRHLLVLQLALLP.....VISFLGLGLGVACIARTR 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 125 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2032	87.4	458	1	CD4_HUMAN
2	2008	86.4	458	1	CD4_PANTR
3	1858	79.9	458	1	CD4_MACFU
4	1857	79.9	458	1	CD4_MACMU
5	1854	79.7	458	1	CD4_MACFA
6	1849	79.5	458	1	CD4_MACNE
7	1832	78.8	458	1	CD4_CERAE
8	1718	74.3	397	1	CD4_ERYPA
9	1718	73.9	397	1	CD4_CERTO
10	1582.5	68.1	457	1	CD4_SAIISC
11	1147	49.3	459	1	CD4_RABIT
12	1137	48.9	463	1	CD4_CANFA
13	999	43.0	457	1	CD4_RAT
14	993	42.7	457	1	CD4_MOUSE
15	293	12.6	240	1	CD7_HUMAN
16	162	7.0	739	1	VCAL_HUMAN
17	155	6.7	739	1	VCAL_HUMAN
18	145	6.2	761	1	NCA2_HUMAN
19	144	6.2	739	1	VCAL_MOUSE
20	141.5	6.1	1493	1	NEOI_MOUSE
21	141	6.1	6885	1	SNE2_HUMAN
22	139	6.0	1447	1	DCR_MOUSE
23	138	5.9	837	1	NCA2_MOUSE
24	137.5	5.9	702	1	CEA5_MOUSE
25	135	5.8	1259	1	SHS1_RAT
26	134.5	5.8	509	1	CAML_RAT
27	134.5	5.8	1260	1	CAML_MOUSE
28	131	5.6	1257	1	CAML_HUMAN
29	130	5.6	1906	1	KMIS_CHICK
30	127	5.5	3707	1	PGM2_MOUSE
31	125.5	5.4	725	1	NCA2_MOUSE
32	125.5	5.4	1447	1	DCC_HUMAN
33	125.5	5.4	4391	1	PGM1_HUMAN

34	125	5.4	349	1	LACH_SCHAM	Q26474 schiocerc
35	125	5.4	2012	1	DSGA_HUMAN	O60469 homo sapien
36	124	5.3	1115	1	NCA1_MOUSE	P13596 mus musculu
37	124	5.3	1298	1	YTFN_HAELN	Q57523 haemophilus
38	123	5.3	837	1	NCA2_HUMAN	O15394 homo sapien
39	122.5	5.3	564	1	C166_BRARE	Q90460 brachydantio
40	122.5	5.3	848	1	NCA1_HUMAN	P13591 homo sapien
41	121.5	5.2	333	1	AMAL_DROME	P15364 drosophila
42	121.5	5.2	338	1	LAMP_HUMAN	Q13449 homo sapien
43	121	5.2	348	1	KILO_RAT	Q92078 ratu
44	121	5.2	997	1	SCPI_RAT	Q03410 ratu
45	120.5	5.2	773	1	PGR_RABIT	P01832 oryctolagus
46	120.5	5.2	1091	1	NCA1_CHICK	P13590 gallus galli
47	120.5	5.2	3305	1	APLP_MANSE	O25490 manduca sex
48	118.5	5.1	740	1	PECI_PIG	O95242 sus scrofa
49	117.5	5.1	338	1	LAMP_RAT	Q62813 ratu
50	117.5	5.1	858	1	NCA1_RAT	P13596 ratu
51	117.5	5.1	1197	1	CAML_BRARE	Q90478 brachydantio
52	116	5.0	555	1	C166_CARAU	Q90304 carassius a
53	116	5.0	853	1	NCA1_BOVIN	P13816 bos taurus
54	115.5	5.0	879	1	PRRP_RAT	Q62786 ratu
55	115.5	5.0	1136	1	TEI1_BOVIN	Q06805 bos taurus
56	115.5	5.0	1302	1	NRG_DROME	P20241 drosophila
57	114	4.9	257	1	FCBA_HUMAN	P12319 homo sapien
58	113.5	4.9	245	1	FCBI_RAT	P12371 ratu
59	113.5	4.9	338	1	LAMP_CHICK	Q98919 gallus galli
60	113.5	4.9	1277	1	CAML_FUGRU	Q98902 tugu rubrip
61	113.5	4.9	1849	1	IGA4_HAELN	P45386 haemophilus
62	113	4.9	1333	1	VGR1_MOUSE	P35969 mus musculu
63	112.5	4.8	359	1	LACH_DROME	Q24372 drosophila
64	112	4.8	588	1	C166_CHICK	P42292 gallus galli
65	111.5	4.8	393	1	ILIS_CERAE	Q29612 cercopithec
66	111.5	4.8	981	1	SCA4_RICE	Q8437 rickettsia
67	110.5	4.8	879	1	PRRP_MOUSE	Q9491 mus musculu
68	110	4.7	777	1	RNI1_RAT	Q64548 ratu
69	110	4.7	1679	1	GCC2_MOUSE	O087g3 mus musculu
70	110	4.7	6632	1	UN89_CAELB	O01761 caenorhabdi
71	109	4.7	1367	1	VGR2_MOUSE	P35918 mus musculu
72	108	4.6	538	1	PVR2_HUMAN	Q22692 homo sapien
73	108	4.6	697	1	STLA_HUMAN	Q96167 homo sapien
74	107.5	4.6	1138	1	TEI1_HUMAN	P35590 homo sapien
75	107.5	4.6	1443	1	NEOI_CHICK	Q90610 gallus galli
76	107	4.6	514	1	CSA_DICDI	P08796 dictyosteli
77	107	4.6	521	1	CEA1_MOUSE	P18809 mus musculu
78	106.5	4.6	1022	1	SCA4_RICPR	Q92d49 rickettsia
79	106	4.6	345	1	OPCM_BOVIN	P11834 bos taurus
80	105.5	4.5	916	1	PERT_BORBR	Q03035 bordetella
81	105.5	4.5	1150	1	ZACA_HUMAN	O06190 homo sapien
82	105	4.5	327	1	112B_BOVIN	P46282 bos taurus
83	105	4.5	345	1	OPCM_HUMAN	Q14982 homo sapien
84	105	4.5	1373	1	RPOB_RICMA	Q9rh43 rickettsia
85	104.5	4.5	515	1	ENV_BIVAV	P25057 bovine leuk
86	104.5	4.5	519	1	ECTO_RAT	P16573 ratu
87	104.5	4.5	1070	1	PRK7_HUMAN	Q13308 homo sapien
88	104	4.5	437	1	E2F2_HUMAN	Q14209 homo sapien
89	104	4.5	524	1	BUTY_MOUSE	Q62556 mus musculu
90	104	4.5	862	1	CD22_MOUSE	P35329 mus musculu
91	104	4.5	1284	1	NRCA_CHICK	P35331 gallus galli
92	104	4.5	1348	1	VGR2_COTJA	P52583 coturnix co
93	104	4.5	1461	1	NEOI_HUMAN	Q92859 homo sapien
94	104	4.5	3210	1	CENF_HUMAN	P49454 homo sapien
95	103.5	4.5	108	1	KYIS_MOUSE	P01652 mus musculu
96	103.5	4.5	129	1	KYIW_HUMAN	P04431 homo sapien
97	103.5	4.5	344	1	NTRI_MOUSE	Q99d10 mus musculu
98	103.5	4.5	344	1	NTRI_RAT	O62782 ratu
99	103.5	4.5	519	1	GAG_SIVAT	P05892 simian immu
100	103.5	4.5	738	1	PECI_HUMAN	P14784 gallus galli
101	103.5	4.5	1010	1	CONT_CHICK	O02815 ovis aries
102	103	4.4	327	1	112B_SHEEP	P35672 salmone
103	103	4.4	562	1	INVG_SALTY	P33459 caprine art
104	103	4.4	1109	1	POL_CAELV	Q9rh41 rickettsia
105	103	4.4	1373	1	RPOB_RICCN	P25504 bovine leuk
106	102.5	4.4	515	1	ENV_BIVAV	

107	102.5	4.4	515	1	ENV BLVJ	P03380 bovine leuk
108	102.5	4.4	521	1	LAG3 MOUSE	O61790 mus musculu
109	102.5	4.4	1097	1	PGDR RAT	O05030 rattus norv
110	102.5	4.4	1134	1	TIE1 MOUSE	O06806 mus musculu
111	102	4.4	508	1	YJ12 YEAST	Q12303 saccharomyc
112	102	4.4	814	1	SLA1_BACAA	P49051 bacillus an
113	102	4.4	992	1	SCA4_RICRY	O9A163 rickettsia
114	102	4.4	1343	1	VGR2 RAT	O08775 rattus norv
115	101.5	4.4	328	1	112B MACMU	P48095 macaca mula
116	101.5	4.4	344	1	NTRI_HUMAN	O9P121 homo sapien
117	101	4.3	135	1	YOR6_ADEB1	P20748 avian adeno
118	101	4.3	1036	1	AVOI_CHICK	P28685 gallus gall
119	101	4.3	1541	1	IGAI_HAEIN	P42782 haemophilus
120	101	4.3	1694	1	SN_MOUSE	O62230 mus musculu
121	101	4.3	2029	1	LAR DROME	P16621 drosophila
122	100.5	4.3	108	1	KV5Q_MOUSE	P01650 mus musculu
123	100.5	4.3	597	1	STIL_PANTR	O95110 pan troglod
124	100.5	4.3	822	1	FGRI_HUMAN	P11362 homo sapien
125	100	4.3	922	1	PERT_BORPA	P43328 bordetella

ALIGNMENTS

RESULT 1
CD4_HUMAN STANDARD; PRT; 458 AA.

AC P01730;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen T4/Leu-3).
GN CD4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85254948; PubMed=2990730;
RA Maddon P.J., Littman D.R., Godfrey M., Maddon D.E., Chess L., Axel R.;
RT "The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: a new member of the immunoglobulin gene family.";
RL Cell 42:93-104(1985).
RN [2]
RP REVISION TO 26.
RX MEDLINE=89028665; PubMed=3263213;
RA Littman D.R., Maddon P.J., Axel R.;
RT "Corrected CD4 sequence.";
RL Cell 55:541-541(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=96303695; PubMed=8723724;
RA Ananari-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S., Malley T., Gibbs R.A.;
RT "A gene-rich cluster between the CD4 and triosephosphate isomerase genes at human chromosome 12p13.";
RL Genome Res. 6:314-326(1996).
RN [4]
RP SEQUENCE FROM N.A., AND VARIANT TRP-265.
RX MEDLINE=91216786; PubMed=1708753;
RA Hodge T.W., Sasse D.R., McDougal J.S.;
RT "Humans with OKT8-epitope deficiency have a single nucleotide base change in the CD4 gene, resulting in substitution of TRP240 for ARG240.";
RL Hum. Immunol. 30:99-104(1991).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS; PubMed=12477932;
RX MEDLINE=22386257;
RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M., Ueda T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Lottoliano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.S., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huljk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Heltón E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywicki M.I., Skalka U., Smillie D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the cellular receptor for simian immunodeficiency virus/human immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
RN [7]
RP SEQUENCE OF 26-394.
RX MEDLINE=90078232; PubMed=2592374;
RA Carr S.A., Hemling M.B., Folsen-Wasserman G., Sweet R.W., Anumula K., Barr J.R., Huddleston M.J., Taylor P.;
RT "Protein and carbohydrate structural analysis of a recombinant soluble CD4 receptor by mass spectrometry.";
RL J. Biol. Chem. 264:21286-21295(1989).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 26-208.
RX MEDLINE=91061881; PubMed=1701030;
RA Wang J., Yan Y., Garrett T.P., Liu J., Rodgers D.W., Garlick R.L., Tarr G.E., Hsain Y., Reinherz E.L., Harrison S.C.;
RT "Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains.";
RL Nature 348:411-418(1990).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 26-208.
RX MEDLINE=91061882; PubMed=2247146;
RA Ryu S.-E., Kwong P.D., Truneh A., Porter T.G., Arthos J., Rosenberg M., Dai X., Xiong N.-H., Axel R., Sweet R.W., Hendrickson W.A.;
RT "Crystal structure of an HIV-binding recombinant fragment of human CD4.";
RL Nature 348:419-426(1990).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (3.9 ANGSTROMS) OF 26-388.
RX MEDLINE=97311402; PubMed=9168119;
RA Wu H., Kwong P.D., Hendrickson W.A.;
RT "dimeric association and segmental variability in the structure of human CD4.";
RL Nature 387:527-530(1997).
RN [11]
RP PALMITOYLATION.
RX MEDLINE=92317088; PubMed=1618861;
RA Crise B., Rose J.K.;
RT "Identification of palmitoylation sites on CD4, the human immunodeficiency virus receptor.";
RL J. Biol. Chem. 267:13593-13597(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

```

CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD4 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd4.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M12807; AAA5572.1; -
DR EMBL; U47924; AAB51309.1; -
DR EMBL; M35160; AAL16069.1; -
DR EMBL; BC025782; AAB25782.1; -
DR PIR; A90872; KMH074.
DR PDB; 1CDH; 30-APR-94.
DR PDB; 1CDI; 30-APR-94.
DR PDB; 3CD4; 31-OCT-93.
DR PDB; 1CDJ; 01-APR-97.
DR PDB; 1CDU; 01-APR-97.
DR PDB; 1CDY; 01-APR-97.
DR PDB; 1WBR; 12-MAR-97.
DR PDB; 1WIO; 07-JUL-97.
DR PDB; 1WIP; 07-JUL-97.
DR PDB; 1WIO; 07-JUL-97.
DR PDB; 1G9M; 27-DEC-00.
DR PDB; 1G9N; 27-DEC-00.
DR PDB; 1GCL; 19-AUG-98.
DR PDB; 1JL4; 19-SEP-01.
DR GLYCOSULEDB; P01730; -
DR Genem; HGNC:1678; CD4.
DR MIM; 166940; -
DR GO; GO:0042101; C:T-cell receptor complex; NAS.
DR GO; GO:0015026; F:coreceptor activity; NAS.
DR GO; GO:0015029; F:internalization receptor activity; TAS.
DR GO; GO:0042289; F:MHC class II protein binding; NAS.
DR GO; GO:0004888; F:transmembrane receptor activity; TAS.
DR GO; GO:0006955; F:immune response; NAS.
DR GO; GO:0009405; F:pathogenesis; TAS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . . NAS.
DR GO; GO:0030217; P:T-cell differentiation; NAS.
DR GO; GO:0045058; P:T-cell selection; NAS.
DR GO; GO:0007168; P:transmembrane receptor protein tyrosine kin. . . NAS.
DR InterPro; IPR000973; CD4_TcRg.
DR InterPro; IPR007110; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PRO0692; CD4TCANTIGEN.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure;
KW Polymorphism.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 374
FT CARBOHYD 296 296
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT LIPID 328 370
FT LIPID 419 419
FT LIPID 422 422
FT VARIANT 265 265
  R -> W (in OKT4-negative populations).

```

```

FT FT STRAND 27 32 /FTid=VAR_003906.
FT FT TURN 33 34
FT FT STRAND 37 39
FT FT TURN 44 45
FT FT STRAND 51 55
FT FT TURN 56 57

Query Match 87.4%; Score 2032; DB 1; Length 458;
Best Local Similarity 89.2%; Pred. No. 6,2e-132;
Matches 405; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

OY 1 MNRGVPFRHLVLYQLALPPATQGNKYVLAQKGDYELCTASQKSIQPHWNSNQIK 60
DB 1 MNRGVPFRHLVLYQLALPPATQGNKYVLAQKGDYELCTASQKSIQPHWNSNQIK 60
OY 61 ILNQGSFLTKGPKSLNDRADRSRLSDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQL 120
DB 61 ILNQGSFLTKGPKSLNDRADRSRLSDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQL 120
OY 121 LVFGLTANSPTHLIQGSLTITLSPGSSPSVCCSPRGKNIQGGKTLVSQLELDG 180
DB 121 LVFGLTANSPTHLIQGSLTITLSPGSSPSVCCSPRGKNIQGGKTLVSQLELDG 180
OY 181 TWTCVLOKOKVEFKIDIVLAFQKASIVYKKEGSEVSPFLAFTVEKLTGSGELMW 240
DB 181 TWTCVLOKOKVEFKIDIVLAFQKASIVYKKEGSEVSPFLAFTVEKLTGSGELMW 240
OY 241 QABRASSKSWITFDLNKKEVSVRVTQDPKLGKGLPHLTLPOALPYAGSGNLTLA 300
DB 241 QABRASSKSWITFDLNKKEVSVRVTQDPKLGKGLPHLTLPOALPYAGSGNLTLA 300
OY 301 LEATGKGLHDEVNLVWNRATQLOKNTCEVWGSPSPKLMSTLKENNEAKSKREKVPVW 360
DB 301 LEATGKGLHDEVNLVWNRATQLOKNTCEVWGSPSPKLMSTLKENNEAKSKREKVPVW 360
OY 361 LNPAGMOCILSDSGOVLSESNIKVLPWTSTPHRPAASALPAPPTGSALPDPTASALP 420
DB 361 LNPAGMOCILSDSGOVLSESNIKVLPWTSTPHRPAASALPAPPTGSALPDPTASALP 420
OY 421 DPRAASALPALAVISFLGLGV-ACVLARTR 453
DB 397 -----MALIVGAGVALLFLGLGIFFCVRCRHR 425

RESULT 2
CD4_PANTR STANDARD; PRT; 458 AA.
AC P16004;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90182664; PubMed=2107024;
RA Camerini D.; Seed B.;
RT "A CD4 domain important for HIV-mediated syncytium formation lies
RT outside the virus binding site.";
RL Cell 60:747-754(1990).
RN [2]
RP SEQUENCE OF 26-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A.; Hirsch V.M.; Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human

```

```

RT Immunodeficiency virus.;
RL Eur. J. Immunol. 22:2973-2981 (1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M31135; AAA35407.1; -.
DR EMBL, X73323; CAAS1749.1; -.
DR PIR, B32722; RMC2T4.
DR HSSP, P01730; 1MIO.
DR GO, GO:0042101; C:T-cell receptor complex; ISS.
DR GO, GO:0015026; F:coreceptor activity; ISS.
DR GO, GO:0042289; F:MHC class II protein binding; ISS.
DR GO, GO:0006955; P:immune response; ISS.
DR GO, GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO, GO:0030217; P:T-cell differentiation; ISS.
DR GO, GO:0045058; P:T-cell selection; ISS.
DR GO, GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro: IPR000973; CD4_TCRg.
DR InterPro: IPR007110; Ig_Like.
DR Pfam, PF00047; Ig_2.
DR PRINTS, PR00692; CD4TCANTIGEN.
DR SMART, SM00406; IGV. 1.
DR PROSITE, PS50835; IG_Like; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396 T-CELL SURFACE GLYCOPROTEIN CD4.
FT TRANSMEM 397 418 POTENTIAL.
FT DOMAIN 419 458 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 126 203 IG-LIKE V-TYPE.
FT DOMAIN 204 317 IG-LIKE C2-TYPE 1.
FT DOMAIN 318 374 IG-LIKE C2-TYPE 2.
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 419 419 S-palmitoyl cysteine (By similarity).
FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
FT CONFLICT 42 62 L -> N (IN REF. 2).
FT CONFLICT 62 62 L -> N (IN REF. 2).
FT CONFLICT 191 191 K -> N (IN REF. 2).
SO SEQUENCE 458 AA; 51057 MW; ATG3AC8A5257D3AD CMC64;
Query Match 86.4%; Score 2008; DB 1; Length 458;
Best Local Similarity 88.3%; Pred. No. 2,7e-130;
Matches 401; Conservative 2; Mismatches 21; Indels 30; Gaps 2;

```

```

Db 121 LVFGLTANSDTHLLQGSLTLTLESPPGSSPVQCSPKQNIQGGKTLVSQLELQDSG 180
Qy 181 TMTCTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGQVFFSPPLAFTVEKLTGSGGLMW 240
Db 181 TMTCTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGQVFFSPPLAFTVEKLTGSGGLMW 240
Qy 241 QAERASSSKSWITFDLKNKEVSVKRYTQDKLQMGKKLPHLTLPOALPOYAGSGNLTLA 300
Db 241 QAERASSSKSWITFDLKNKEVSVKRYTQDKLQMGKKLPHLTLPOALPOYAGSGNLTLA 300
Qy 301 LEAKTGLHGEVNLVVMRATQLOKNLTCEWGTSPPLMLSLKLNKAAYSKREKVMV 360
Db 301 LEAKTGLHGEVNLVVMRATQLOKNLTCEWGTSPPLMLSLKLNKAAYSKREKVMV 360
Qy 361 LNEBAGMOCGLSDSGVGLSENIKVLPTWSTPVHPPASALPAPPTGSALPDQTASALP 420
Db 361 LNEBAGMOCGLSDSGVGLSENIKVLPTWSTPVHPPASALPAPPTGSALPDQTASALP 420
Qy 421 DPPAASLPALAVISFLGLGV-ACVLAARR 453
Db 397 -----MALIVLGVAAGLLPFGIGIFCVCRHR 425

RESULT 3
CD4_MACFU STANDARD; PRT; 458 AA.
ID CD4_MACFU
AC P79184;
DT 15-JUN-1998 (Rel. 36, Created)
DT 15-JUN-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
GN CD4.
OS Macaca fuscata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9543;
RN [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tatsumi M.;
RL Submitted (FEBS-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, D63348; BAA09672.1; -.
DR HSSP, P01730; 1MIR.
DR GO, GO:0042101; C:T-cell receptor complex; ISS.
DR GO, GO:0015026; F:coreceptor activity; ISS.
DR GO, GO:0042289; F:MHC class II protein binding; ISS.
DR GO, GO:0006955; P:immune response; ISS.
DR GO, GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO, GO:0030217; P:T-cell differentiation; ISS.
DR GO, GO:0045058; P:T-cell selection; ISS.
DR InterPro: IPR000973; CD4_TCRg.
DR InterPro: IPR007110; Ig_Like.
DR Pfam, PF00047; Ig_2.
DR PRINTS, PR00692; CD4TCANTIGEN.
DR SMART, SM00406; IGV. 1.

```



```

DR PROSITE; PS50835; IG_Like; 1.
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT CAROXYD 42 42
FT CAROXYD 26 296
FT CAROXYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 422
SQ SEQUENCE 458 AA; 50828 MW; 76B3E7E08185535 CRC64;

Query Match          79.9%; Score 1858; DB 1; Length 458;
Best Local Similarity 81.3%; Pred. No. 5,1e-120;
Matches 369; Conservative 18; Mismatches 37; Indels 30; Gaps 2;

QY 1 MNRGVPFRLLLVLTALPAATQGNKVYLGKKGTVELTCTASQKSIQFMKNSNOIK 60
DB 1 MNRGVPFRLLLVLTALPAATQGNKVYLGKKGTVELTCTASQKSIQFMKNSNOIK 60
QY 61 ILGNQSFLLTKGPSKLNDRADSRSLMDQGNPLIIKULKIDSDTYICEVDQKEEYOL 120
DB 61 ILGNQSFLLTKGPSKLNDRADSRSLMDQGNPLIIKULKIDSDTYICEVDQKEEYOL 120
QY 121 LVFGLTANSDFHLLQGSITLTLLESPGSSPSVQCRSPRKNIIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDFHLLQGSITLTLLESPGSSPSVQCRSPRKNIIQGGKTLVSQLELDQSG 180
QY 121 LVFGLTANSDFHLLQGSITLTLLESPGSSPSVQCRSPRKNIIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDFHLLQGSITLTLLESPGSSPSVQCRSPRKNIIQGGKTLVSQLELDQSG 180
QY 161 TWTCTVLOQKQKVEKIDIVLAFQASSTIVYKKEGEQVSEFPPLAFVEKLTSGGELMW 240
DB 161 TWTCTVLOQKQKVEKIDIVLAFQASSTIVYKKEGEQVSEFPPLAFVEKLTSGGELMW 240
QY 161 TWTCTVLOQKQKVEKIDIVLAFQASSTIVYKKEGEQVSEFPPLAFVEKLTSGGELMW 240
DB 161 TWTCTVLOQKQKVEKIDIVLAFQASSTIVYKKEGEQVSEFPPLAFVEKLTSGGELMW 240
QY 241 QAEERASSSSKSWITTPOLKKEVSVKRVTOPKLOMGKKPLHLLTLPQALPOVAGSNNLTLA 300
DB 241 QAEERASSSSKSWITTPOLKKEVSVKRVTOPKLOMGKKPLHLLTLPQALPOVAGSNNLTLA 300
QY 301 LEAKTGKTLHGVNVLVVRATQLOKNLTCEVWGPTSPKMLSLKLNKEAKYSKREKPYVW 360
DB 301 LEAKTGKTLHGVNVLVVRATQLOKNLTCEVWGPTSPKMLSLKLNKEAKYSKREKPYVW 360
QY 301 LEAKTGKTLHGVNVLVVRATQLOKNLTCEVWGPTSPKMLSLKLNKEAKYSKREKPYVW 360
DB 301 LEAKTGKTLHGVNVLVVRATQLOKNLTCEVWGPTSPKMLSLKLNKEAKYSKREKPYVW 360
QY 361 LNPEAGMMOCLLSDSGOVLLESNIVLPTWSTPVHPRASALPAPPTGSALPDPTASALP 420
DB 361 LNPEAGMMOCLLSDSGOVLLESNIVLPTWSTPVHPRASALPAPPTGSALPDPTASALP 420
QY 421 DEPPASALPALAVISPLIGLIGV-ACYLART 453
DB 421 DEPPASALPALAVISPLIGLIGV-ACYLART 453
QY 397 -----MALIVLGAGVAGLLFTGLGIFPCVCRHR 425
DB 397 -----MALIVLGAGVAGLLFTGLGIFPCVCRHR 425

RESULT 4
CD4_MACMU STANDARD; PRT; 458 AA.
AC P16T03; Q29617;
DT 01-APR-1990 (Rel. 14, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Leu-3).
GN CD4.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopitheciinae; Macaca.

```

```

OX NCBI_TaxID=9544;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=90182664; PubMed=2107024;
RA Cameroni D., Seed B.;
RT "A CD4 domain important for HIV-mediated syncytium formation lies
RL outside the virus binding site.";
RN [2]
RN [3]
RN [4]
RN SEQUENCE FROM N.A.
RN TISSUE=Thymocytes;
RA Hashimoto O., Tatsumi M.;
RT "Molecular cloning and expression of macaque CD4s.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE OF 28-424 FROM N.A.
RN TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
RN [6]
RN SEQUENCE OF 107-192 FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.B., DiBotto T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (Primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
RN [7]
RN FUNCTION: Accessory protein for MHC class-II antigen/T-cell
RN receptor interaction. May regulate T-cell activation.
RN SUBUNIT: Associates with p56-lck (By similarity).
RN SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
RN SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.ebi.ac.uk/announcements
RN or send an email to license@ebi.ac.uk).
RN EMBL; M31134; AA36838.1; -
RN EMBL; D63347; BAA09671.1; -
RN EMBL; X73326; CAA51752.1; -
RN EMBL; AF057385; AAC5129.1; -
RN HSPB, P01730; IMR.
RN GO; GO:0042101; C: T-cell receptor complex; ISS.
RN GO; GO:0015026; P: coreceptor activity; ISS.
RN GO; GO:0042289; F: MHC class II protein binding; ISS.
RN GO; GO:0006955; P: immune response; ISS.
RN GO; GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
RN GO; GO:0030217; P: T-cell differentiation; ISS.
RN GO; GO:0045058; P: T-cell selection; ISS.
RN GO; GO:0007169; P: transmembrane receptor protein tyrosine kin. . .; ISS.
RN InterPro: IPR000973; CD4_TCSG.
RN InterPro: IPR007110; IG-like.
RN InterPro: IPR003596; IG_V.
RN Pfam; PF00047; Ig_2.
RN PRINTS; PR00692; CD4TCANTGEN.
RN SMART; SM00406; IGV_1.
RN PROSITE; PS50835; IG_Like; 1.
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT CAROXYD 42 42
FT CAROXYD 26 296
FT CAROXYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 422
SQ SEQUENCE 458 AA; 50828 MW; 76B3E7E08185535 CRC64;

```

```

FT DOMAIN 126 203 IG-LIKE C2-TYPE 1.
FT DOMAIN 204 317 IG-LIKE C2-TYPE 2.
FT DOMAIN 318 374 IG-LIKE C2-TYPE 3.
FT CARBOHYD 296 326 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 419 419 S-palmitoyl cysteine (By similarity).
FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
FT CONFLICT 42 42 N -> T (IN REF. 1).
FT CONFLICT 62 62 L -> S (IN REF. 3).
FT CONFLICT 67 67 L -> S (IN REF. 2).
FT CONFLICT 169 169 I -> L (IN REF. 2).
FT CONFLICT 191 191 K -> N (IN REF. 3).
FT CONFLICT 248 248 S -> P (IN REF. 2).
FT CONFLICT 265 265 R -> Q (IN REF. 3).
FT CONFLICT 349 349 A -> T (IN REF. 2).
SQ SEQUENCE 458 AA; 50884 MW; 8BB80339FAFE0808 CRC64;

```

Query Match 79.9%; Score 1857; DB 1; Length 458;

Best Local Similarity 81.3%; Pred. No. 6e-120; Mismatches 37; Indels 30; Gaps 2;

```

Matches 369; Conservative 18; Mismatches 37; Indels 30; Gaps 2;

QY 1 MNRGVPFRHLLVQLALPPATQGNKVVLGKKGDTVELTCTASQKKSIOFHMKNNSQIK 60
DB 1 MNRGIPFRHLLVQLALPPATQGNKVVLGKKGDTVELTCTASQKKNTOFHMKNNSQIK 60
QY 61 ILNGQSFLLTKGSKLNDRAKRSRLMDQGNFPLIKLKIEDSDTYICEVENEKEVEL 120
DB 61 ILGIGFLLTKGSKLNDRAKRSRLMDQGNFPLIKLKIEDSDTYICEVENEKEVEL 120
QY 121 LVFGILTANSDTHLLEGGSLTLTLESPGSSPSVOCSPRGKNIQGGKTLVSQLELDSG 180
DB 121 LVFGILTANSDTHLLEGGSLTLTLESPGSSPSVOCSPRGKNIQGGKTLVSQLELDSG 180
QY 121 LVFGILTANSDTHLLEGGSLTLTLESPGSSPSVOCSPRGKNIQGGKTLVSQLELDSG 180
DB 121 LVFGILTANSDTHLLEGGSLTLTLESPGSSPSVOCSPRGKNIQGGKTLVSQLELDSG 180
QY 181 TWCTCTVLOMOKKVEFKIDIVLAFQKASSIYKKEGEVEFSPLAFTVEKLTGSGELMW 240
DB 181 TWCTCTVLOMOKKVEFKIDIVLAFQKASSIYKKEGEVEFSPLAFTVEKLTGSGELMW 240
QY 241 QAEERASSKSWITPDLKKEVSVKRVNODPKLOMGKTLPLTLTPQALPOYAGSGNLTLA 300
DB 241 QAEERASSKSWITPDLKKEVSVKRVNODPKLOMGKTLPLTLTPQALPOYAGSGNLTLA 300
QY 301 LEAKTGKLGHOENVLVWRATOLQKALICEVWGPTSPKMLSLKLENKAKYKKEKPYWV 360
DB 301 LEAKTGKLGHOENVLVWRATOLQKALICEVWGPTSPKMLSLKLENKATYSKQAKAVWV 360
QY 361 LNPEAGMOCILSDSGOVLLESNIKVLTFTWSTPVHPRASALPAPPTGSLPDPQTASALP 420
DB 361 LNPEAGMOCILSDSGOVLLESNIKVLTFTWSTPVHPRASALPAPPTGSLPDPQTASALP 420
QY 421 DPPASALPALVALVIFLGLGLG-ACVLARTR 453
DB 397 -----MALIVGVAAGLLFTGLGIFFCVRCRHR 425

```

RESULT 5

CD4_MACFA STANDARD; PRT; 458 AA.

```

AC P79185;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
   T4/Leu-3).
GN CD4.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
NCBI_TaxId=9541;
OX
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=Thymocytes;
RA Tatsumi M., Yabe M., Yamada Y.K.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -!- SUBUNIT: Associates with p56-lck (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

```

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

DR EMBL; D63349; BAA09673.1; -.
DR HSSP; P01730; IWR.
DR GO; GO:0042101; C-T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4-TCRg.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig; v.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; ISY; 1.
DR PROSITE; PS50835; IG_LIKE_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSLEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 396
FT DOMAIN 419 458
FT DOMAIN 126 203
FT DOMAIN 204 317
FT CARBOHYD 42 42
FT CARBOHYD 296 326
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
SQ SEQUENCE 458 AA; 50874 MW; 9105479FB5C56FF7 CRC64;

```

Query Match 79.7%; Score 1854; DB 1; Length 458;

Best Local Similarity 81.3%; Pred. No. 9.6e-120; Mismatches 38; Indels 30; Gaps 2;

```

Matches 369; Conservative 17; Mismatches 38; Indels 30; Gaps 2;

QY 1 MNRGVPFRHLLVQLALPPATQGNKVVLGKKGDTVELTCTASQKKSIOFHMKNNSQIK 60
DB 1 MNRGIPFRHLLVQLALPPATQGNKVVLGKKGDTVELTCTASQKKNTOFHMKNNSQIK 60
QY 61 ILNGQSFLLTKGSKLNDRAKRSRLMDQGNFPLIKLKIEDSDTYICEVENEKEVEL 120
DB 61 ILGIGFLLTKGSKLNDRAKRSRLMDQGNFPLIKLKIEDSDTYICEVENEKEVEL 120
QY 121 LVFGILTANSDTHLLEGGSLTLTLESPGSSPSVOCSPRGKNIQGGKTLVSQLELDSG 180
DB 121 LVFGILTANSDTHLLEGGSLTLTLESPGSSPSVOCSPRGKNIQGGKTLVSQLELDSG 180

```

Oy	181	TWTCGTVLONOKKVEEKIDIVLAPOKASISVYKKGEQVESFPALFTPEKLTGSGLMW	240
Dd	181	TWTCVSODQKTVEREKIDIIVLAFKASSTVYKKGEQVESFPALFTLEKLTGSGLMW	240
Oy	241	QAERASSSSKSMTTPELKNKEVSVKRWTDPPKLQMKQLPLHLTLTPQALPQYAGSNLTIA	300
Dd	241	QAERASSSSKSMTTFPLLKNKEVSVKRWTDPPKLQMKQLPLHLTLTPQALPQYAGSNLTIA	300
Oy	301	LEAKTGKHQEVNLVVMRATOLQNKLTCEVWGPTSPKIMLSIKLENKEAKYSKREKPVVV	360
Dd	301	LEAKTGKHQEVNLVVMRATOPQENKLTCEVWGPTSPKLTLSIKLENKGTWSKAQAVVV	360
Oy	361	LNPEGMQMOCLLSDSGOVULBSEINIVLPWPSFPHVPNPRASALPAPRTGSALPDPPQTASALP	420
Dd	361	LNPEGMQMOCCLSDSGOVULBSEINIKVPIWPVP-----	396
Oy	421	DPNAPSALPALAVISFLIGLGSLG-ACLTARTR	453
Dd	397	-----MALIVGGVAGLLFLFGSLGFPCVRCNHR	425

RESULT 6
CD4_MACNE
ID CD4_MACNE STANDARD PRT 458 AA.

```

AC   Q08J40; P79196;
DT    01-FEB-1995 (Rel. 31, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE    T-cell surface glycoprotein Cd4 precursor (T-cell surface antigen
      T4/Lew-3).
OS    Cd4.
CN    Macaca nemestrina (pig-tailed macaque).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC    Cercopithecinae; Macaca.
OX    NCBI_Taxid=9545;
RN     [1]
RP     SEQUENCE FROM N.A.
RA     Hashimoto O., Tatsumi M.;
RL     Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN     [2]
RP     SEQUENCE OF 28-424 FROM N.A.
RC     TISSUE=Blood;
RX     MEDLINE=93049640; PubMed=1425921;
RA     Fomsgaard A., Hirsch V.M., Johnson P.R.;
RL     "Cloning and sequences of primate CD4 molecules: diversity of the
      cellular receptor for simian immunodeficiency virus/human
      immunodeficiency virus.";
      Eur. J. Immunol. 22:2973-2981(1992).
CC     -1 FUNCTION: Accessory protein for MHC class-II antigen/T-cell
      receptor interaction. May regulate T-cell activation.
CC     -1 SUBUNIT: Associates with p56-lck (By similarity).
CC     -1 SUBCELLULAR LOCATION: Type I membrane protein.
CC     -1 SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC     -1 SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC
CC     This SWISS-PROT entry is copyright. It is produced through a collaboration
CC     between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC     the European Bioinformatics Institute. There are no restrictions on its
CC     use by non-profit institutions as long as its content is in no way
CC     modified and this statement is not removed. Usage by and for commercial
CC     entities requires a license agreement (See http://www.isdb-sib.ch/announce/
CC     or send an email to license@sib-ch).


---



EMBL; D63346; BAA09670.1; -.  
EMBL; X73325; CAAS1751.1; -.  
HSSP; P01730; IMBR.  
DR GO; GO:0042101; C:T-cell receptor complex; ISS.  
DR GO; GO:0015026; F:coreceptor activity; ISS.  
DR GO; GO:0042289; F:MHC class II protein binding; ISS.  
DR GO; GO:0006955; P:immune response; ISS.  
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. .; ISS.  
DR GO; GO:0030217; P:T-cell differentiation; ISS.


```

DR	GO: 0045058; P: T-cell selection; ISS.
DR	GO: 0007716; P: Transmembrane receptor protein tyrosine kin. . .; ISS
DR	InterPro; IPR000973; CD4_TcRg.
DR	InterPro; IPR007110; Ig-like.
DR	InterPro; IPR003596; Ig_v.
DR	Pfam; PF00047; Ig. 2.
DR	PRINTS; PR00692; CD4TCANTIGEN.
DR	SMART; SM00406; IGV: 1.
DR	PROSITE; PS50835; IG_LIKE: 1
KW	Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW	Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT	SIGNAL 1 25
FT	CHAIN 26 458
FT	DOMAIN 26 396
FT	TRANSMEM 397 418
FT	DOMAIN 419 458
FT	DOMAIN 26 125
FT	DOMAIN 126 203
FT	DOMAIN 204 317
FT	DOMAIN 318 374
FT	CARBOHYD 42 42
FT	CARBOHYD 295 296
FT	CARBOHYD 325 325
FT	DISULFID 41 109
FT	DISULFID 155 184
FT	DISULFID 328 370
FT	LIPID 419 419
FT	LIPID 422 422
FT	CONFLICT 57 57
FT	CONFLICT 91 91
FT	CONFLICT 105 105
FT	CONFLICT 113 113
FT	CONFLICT 302 302
FT	CONFLICT 349 349
SO	SEQUENCE 458 AA; 50905 MW; 751a9BA2C8B3EB16 CRC64;

Query Match	79.5%	Score 1849;	DB 1;	Length 458;
Best Local Similarity	80.8%;	Pred. No. 2.1e-119;		
Matches 367;	Conservative 20;	Mismatches 37;	Indels 30;	Gaps 2

```

Oy 1 MNRGVRPRLVLVLQVALLPAATQGNKVLLKRGDYEVLCTTAQOKSIOEHMKNNOIK 60
Db 1 MNRGIPRRLHLLVLQVALLPAVATQGRKVVLLGKKGDEYELTCMAQKKNTPHMKNSDQIK 60
Oy 61 ILNGQSSFLTKGPKSLNDRADSRRLMDQGNFPLIKLAKIEDSDYICVEYDQKEVQL 120
Db 61 ILGIQSSFLTKGPKSLNDRADSRRLMDQGCMSIINKLKIEDSNTYICVEYNEKEVEL 120
Oy 121 LVNGLTANSDTHLLOGOSITLTLEBPSSGSPSVOCRSPRGKNIQGGKTLVSQLELDSC 180
Db 121 LVNGLTANSDTHLLEQOSITLTLEBPSSPSVYKCRSPGKNIQGGKTLSPOLERDSC 180
Oy 181 TWFCTYLOQOKYVEFIDIVLAFQKASSIYYKKRGEVERSPFLAFVEXLTGSGELMW 240
Db 181 TWFCTYSQOKYVEFIDIVLAFQKASSIYYKKRGEVERSPFLATTEKLTGSGELMW 240
Oy 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLOMGKPLHLTLPOALPOYAGSGNLTTLA 300
Db 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLOMGKPLHLTLPOALPOYAGSGNLTTLA 300
Oy 301 LEAKTKLQOEVLVVMRATQLOKULTEWGPPTSPLKMLSKLENKEAVYSRKRKVWY 360
Db 301 LDKATKLEQOEVLVVMRATQFOENLTCEWGPPTSPLKTLTKLENKGTTVSQAARVWY 360
Oy 361 LNEBAGMOCCLSDSGOVLNESINIKVLTWMTSPHPRASALPAPPTGSALPDQOTASALP 420
Db 361 LNEBAGMOCCLSDSGOVLNESINIKVLTWMTSPHPRASALPAPPTGSALPDQOTASALP 420
Oy 421 DPPASALPAAVLAVISFLIGLGLSV-ACVLAATR 453
Db 397 -----MALIVLGGVAGELLFTGLGTFVCVRCHR 425

```

```

RESULT 7
CD4_CERAB STANDARD: PRT: 458 AA.
AC 006338: 002805: 07593: 028217;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
  T4/Leu-3).
NM CD4.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tatsumi M.;
RT "Molecular cloning and expression of african green monkey CD4.",
RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93043640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
  cellular receptor for simian immunodeficiency virus/human
  immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
RN [3]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Peripheral blood;
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Mueller-Trutwin M.C., Diop O., Hansen J., Machiot C.,
  Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
  their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
RN [4]
RP SEQUENCE OF 107-192 FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., DiCoccia T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
  mangabeys (Primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
  receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (by similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D86589; BAAL3132.1; -
DR EMBL; X73322; CA51748.1; -
DR EMBL; AF001226; AAB60873.1; -
DR EMBL; AF001228; AAB60875.1; -
DR EMBL; AF057380; AAC5124.1; -
DR HSPB; P01730; IWO.
DR GO; GO:004210; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.

```

```

DR GO; GO:0007169; P:Transmembrane receptor protein tyrosine kin. .; ISS.
DR InterPro: IPR000973; CD4 TcAg.
DR InterPro: IPR007110; Ig-Like.
DR InterPro: IPR003596; Ig_v.
DR Pfam: PF00047; Ig; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KM Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
  Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSSEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 374
FT CARBOHYD 42 42
FT CARBOHYD 281 281
FT CARBOHYD 296 296
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
FT CONFLICT 46 46
FT CONFLICT 59 59
FT CONFLICT 115 115
FT CONFLICT 165 165
FT CONFLICT 200 200
FT CONFLICT 227 227
FT CONFLICT 271 271
FT CONFLICT 281 281
SQ SEQUENCE 458 AA; 51158 MW; FC523D2DD1F72B7 CRC64;

Query Match 78.8%; Score 1832; DB 1; Length 458;
Best Local Similarity 80.4%; Pred. No. 3,1e-118;
Matches 365; Conservative 18; Mismatches 41; Indels 30; Gaps 2;

QY 1 MNRGVPFRHLVLYQALPEPATQGNKRVLYGKKGDTVELTCTASQKKSIOFHNKNSNOIK 60
DB 1 MNRGVPFRHLVLYQALPEPATQGNKRVLYGKKGDTVELTCTASQKKSIOFHNKNSNOIK 60
QY 61 ILGNQGSFLLTKGSSKLANDRSDRSIMDQGNFLLIKNLKIDSDTYTCVEBQKEVQL 120
DB 61 ILGNQGSFLLTKGSSKLANDRSDRSIMDQGNFLLIKNLKIDSDTYTCVEBQKEVQL 120
QY 61 ILKQGSFLLTKGSSKLRDRIDSRKSLMDQCFSMIIKNLKIETSETYICVENKKEVEL 120
DB 61 ILKQGSFLLTKGSSKLRDRIDSRKSLMDQCFSMIIKNLKIETSETYICVENKKEVEL 120
QY 121 LVFGLTANSTHTLLQGOSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
DB 121 LVFGLTANSTHTLLQGOSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG 180
QY 181 TWTCTVLYQNKVEFKIDIVLAFOKASSIVYKKEGBOVESPLAFTVEKLTGSGELMW 240
DB 181 TWTCTVLYQNKVEFKIDIVLAFOKASSIVYKKEGBOVESPLAFTVEKLTGSGELMW 240
QY 181 TWTCTVLYQNKVEFKIDIVLAFOKASSIVYKKEGBOVESPLAFTVEKLTGSGELMW 240
DB 181 TWTCTVLYQNKVEFKIDIVLAFOKASSIVYKKEGBOVESPLAFTVEKLTGSGELMW 240
QY 241 QABRASSSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPLHLPLPOLPYAASSGNITLA 300
DB 241 QABRASSSKSWITFDLKNKEVSVKRYTQDPKLOMGKKLPLHLPLPOLPYAASSGNITLA 300
QY 301 LEAKTGKLEHGVNLVYMRATOLQKNLTCEYWGTSPLMLSLLENBEAIVSRKEXPVWV 360
DB 301 LEAKTGKLEHGVNLVYMRATOLQKNLTCEYWGTSPLMLSLLENBEAIVSRKEXPVWV 360
QY 361 LNEBAGMOCILSDSQVLLIESNIKYLPWTSTFVHPRASALPAPPTGSALPDQRTASALP 420
DB 361 LNEBAGMOCILSDSQVLLIESNIKYLPWTSTFVHPRASALPAPPTGSALPDQRTASALP 420
QY 421 DPAASALPAAALVIFSLGLGLGV-ACVLARTR 453
DB 421 DPAASALPAAALVIFSLGLGLGV-ACVLARTR 453
QY 397 -----MALIVLGVAGILLFTGLGIFFCVCRHR 425
DB 397 -----MALIVLGVAGILLFTGLGIFFCVCRHR 425

```

```

RESULT 8
CD4_ERYPA STANDARD: PRT; 397 AA.
ID CD4_ERYPA
AC Q08339;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Lew-3)
DE (Fragment).
GN CD4.
OS Erythrocybus patas (Red guenon) (Husar).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Erythrocybus.
OX NCBI_TaxId=9538;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus."
RL Bur. J. Immunol. 22:2973-2981(1992).
CC -!- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -!- SUBUNIT: Associates with p56-lck (By similarity).
CC -!- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X73324; CAA51750.1; -.
DR HSSP; P01730; IMIO.
DR GO; GO:0042101; C/T-cell receptor complex; ISS.
DR GO; GO:0015025; Fc-gamma receptor activity; ISS.
DR GO; GO:0042289; Fc-gamma class II protein binding; ISS.
DR GO; GO:0006955; P-immune response; ISS.
DR GO; GO:0045086; P-positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P-T-cell differentiation; ISS.
DR GO; GO:0045058; P-T-cell selection; ISS.
DR GO; GO:0007169; P-transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_V.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS50835; IG_LIKE_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Lipoprotein; Palmitate.
FT NON TER 1
FT DOMAIN <1 369 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 370 391 POTENTIAL.
FT DOMAIN 392 >397 CYTOPLASMIC (POTENTIAL).
FT DOMAIN <1 98 IG-LIKE V-TYPE.
FT DOMAIN 99 176 IG-LIKE C2-TYPE 1.
FT DOMAIN 177 290 IG-LIKE C2-TYPE 2.
FT DOMAIN 291 347 IG-LIKE C2-TYPE 3.
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 14 82 BY SIMILARITY.
FT DISULFID 128 157 BY SIMILARITY.
FT DISULFID 301 343 BY SIMILARITY.

```

```

FT LIPID 392 392 S-palmitoyl cysteine (By similarity).
FT LIPID 395 395 S-palmitoyl cysteine (By similarity).
FT NON TER 397 397
SQ SEQUENCE 397 AA; 44081 MW; 678879797A6B7BA4F CRC64;

Query Match
Best Local Similarity 80.5%; Pred. No. 3,4e-11;
Matches 342; Conservative 16; Mismatches 37; Indels 30; Gaps 2;

QY 28 VVLGKGGTVLTLCTASQKSIQFHWKNSNQIKILGNGSFLTKPSTLNDRAQSRSLM 87
DB 1 VVLGKGGTVLTLCTASQKTIQFHWKNSNQIKILGKGSFLTKGPSKLRRTTSRKSILM 60
QY 88 DQGNPLILIKMLKEDPSITYCEVEDKEEYQVLLVFGGLTASDPHLLGQSITLTLSP 147
DB 61 DQGCSTMIKMLKEDPSITYCEVEDKEEYVLLVFGGLTASDPHLLGQSITLTLSP 120
QY 148 GSPSPVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLQNKKEFKIDIVLAFQKA 207
DB 121 GSPSPVKCRSPRGKNIQGGRTLSVPQLERQDSGTWCTVSGDQNTVERKIDIVLAFQKA 180
QY 208 SSIVYKKEGEYERSFPLAFVEKLTGSGELMWQAEARSSSKSMTTFLPKNKEVSKVKT 267
DB 191 SSIVYKKEGEYERSFPLAFVEKLTGSGELMWQAEARSSSKSMTTFLPKNKEVSKVKT 240
QY 268 QDPKLQWKGKPLHLTLTPOALPOYAGSGLTLTLEAKTGKHOEYVNLVWMRATOLQKMLT 327
DB 241 QDPKLQWKGKPLHLTLTPOALPHYAGSGLTLTLEAKTGKHOEYVNLVWMRATOLQKMLT 300
QY 328 CEVWGPSTPKMLSLKLENKAKVSKREKPYVNLNPEAGMQLSDSGOVLLESNIVLT 387
DB 301 CEVWGPSTPKMLSLKLENKAKVSKREKPYVNLNPEAGMQLSDSGOVLLESNIVLT 360
QY 388 PTVSPVPRASALPAPPTGSALPDPPTASALPDPPASALPALALVSLGLGLGV-A 446
DB 361 PTVSPVPRVPP-----MALIVLGVGAGLLFTGLGIF 391
QY 447 CVLAR 451
DB 392 CVRCR 396

RESULT 9
CD4_CERTO STANDARD: PRT; 397 AA.
ID CD4_CERTO
AC Q08336;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Lew-3)
DE (Fragment).
GN CD4.
OS Cercopithecus torquatus atys (Red-crowned mangabey) (Sooty mangabey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxId=9531;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus."
RL Bur. J. Immunol. 22:2973-2981(1992).
CC -!- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -!- SUBUNIT: Associates with p56-lck (By similarity).
CC -!- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: X73327; CAAS1754.1; -;
 DR EMBL: X73327; CAAS1753.1; -;
 DR HSPB; P01730; 1M7Q.
 DR GO; GO:0042101; C:T-cell receptor complex; ISS.
 DR GO; GO:0015026; F:coreceptor activity; ISS.
 DR GO; GO:0042289; F:MHC class II protein binding; ISS.
 DR GO; GO:0006955; P:immune response; ISS.
 DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
 DR GO; GO:0030217; P:T-cell selection; ISS.
 DR GO; GO:0007169; P:Transmembrane receptor protein tyrosine kin. . .; ISS.
 DR InterPro; IPR000973; CD4 TCAG.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003596; IG_v.
 DR Pfam; PF00047; Ig; 2.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
 KW Immune response; Repeat; Lipoprotein; Palmitate.
 FT NON_TER 1 1
 FT DOMAIN 1 369
 FT TRANSMEM 370 391
 FT DOMAIN 392 >397
 FT DOMAIN <1 98
 FT DOMAIN 99 176
 FT DOMAIN 177 290
 FT CARBOHYD 291 347
 FT CARBOHYD 15 15
 FT CARBOHYD 269 269
 FT CARBOHYD 298 298
 FT DISULFID 14 82
 FT DISULFID 128 157
 FT DISULFID 301 343
 FT LIPID 392 392
 FT LIPID 395 395
 FT VARIANT 20 20
 FT VARIANT 43 43
 FT VARIANT 86 86
 FT VARIANT 96 96
 FT VARIANT 173 173
 FT VARIANT 316 316
 FT NON_TER 397 397
 SQ SEQUENCE 397 AA; 43926 MW; 8660B636D2DB38A7 CRC64;

Query Match 73.9%; Score 1718; DB 1; Length 397;
 Best Local Similarity 80.2%; Pred. No. 1.6e-110;
 Matches 341; Conservative 19; Mismatches 35; Indels 30; Gaps 2;

QY 28 VILGKGGDTVELTCTASQKSKISQFMKNSKQIKLIGNGSFLTGPSTLNRAASRLM 87
 DB 1 VVLGKGGDTVELTCTASQKSKISQFMKNSKQIKLIGNGSFLTGPSTLNRAASRLM 60
 QY 88 DQGNFLLIKNLKLEDSDTYICEDVQKEEVLVFGLTANSDFHLQGGSLTLTLESP 147
 DB 61 DQGCFTMIKLIKLEDSFTYICENVNKEVELVFGLTANSDFHLQGGSLTLTLESP 120
 QY 148 GSSPEVQCRSPKNGIKGSGKTLVSQELQDSGTCTVLNOKKVEFKIDIVLAFOKA 207
 DB 121 GSSPEVKRSRPGKNGIKGSGKTLVSQELQDSGTCTVLNOKKVEFKIDIVLAFOKA 180
 QY 208 SSIYVKKGEQVEFPLATFWEKLTGSGELMWQKERRSSSKSWTTPCLKKEVSVKRV 267
 DB 181 SSIYVKKGEQVEFPLATFWEKLTGSGELMWQKERRSSSKSWTTPCLKKEVSVKRV 240

QY 268 QDPFLQWGGKLLPHLTLPQALPOYAGSGNTLALFAKTKLHDEVNLVWKRATOLQKNT 327
 DB 241 QDPFLQWGGKLLPHLTLPQALPOYAGSGNTLALFAKTKLHDEVNLVWKRATOLQKNT 300
 QY 328 CEVWGPTSPKLMISLKIENKAKVSKREKPVWVNLNPAQMQLLSDSGVLLSNTKVL 387
 DB 301 CEVWGPTSPKLTLSLLENKAKVSKOKAKVWVNLNPAQMQLLSDSGVLLSNTKIV 360
 QY 388 PTWSTPHRPAASALPAPPTGSALPDPOTASALPDPAPASALPALANVLSFLGLGIGV-A 446
 DB 361 PTWSTPHVOP-----MALIVGVAGLLFTGLGIFP 391
 QY 447 CVLAR 451
 DB 392 CVRCR 396

RESULT 10
 CD4_SAISC STANDARD; PRT; 457 AA.
 ID SAISC
 AC 029037;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
 T4/Leu-3).
 GN CD4.
 OS Saimiri sciureus (Common squirrel monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Saimiri.
 ON NCBI_TaxID=9521;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tatsuaki M., Hashimoto O.;
 RL Submitted (Aug-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
 CC receptor interaction. May regulate T-cell activation.
 CC -1- SUBUNIT: Associates with p56-lck (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: D86588; BAA131.1; -;
 DR HSPB; P01730; 1M7R.
 DR GO; GO:0042101; C:T-cell receptor complex; ISS.
 DR GO; GO:0015026; F:coreceptor activity; ISS.
 DR GO; GO:0042289; F:MHC class II protein binding; ISS.
 DR GO; GO:0006955; P:immune response; ISS.
 DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
 DR GO; GO:0030217; P:T-cell selection; ISS.
 DR GO; GO:0045058; P:T-cell selection; ISS.
 DR InterPro; IPR000973; CD4 TCAG.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003596; IG_v.
 DR Pfam; PF00047; Ig; 2.
 DR PRINTS; PR00692; CD4TCANTIGEN.
 DR SMART; SM00406; IGV; 1.
 DR PROSITE; PS50835; IG_LIKE; 1.
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
 KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
 FT SIGNAL 1 25
 FT CHAIN 26 457
 FT DOMAIN 26 395
 FT TRANSMEM 396 417
 FT POTENTIAL.

```

FT DOMAIN 418 457 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 125 IG-LIKE V-TYPE.
FT DOMAIN 126 202 IG-LIKE C2-TYPE 1.
FT DOMAIN 203 316 IG-LIKE C2-TYPE 2.
FT DOMAIN 317 373 IG-LIKE C2-TYPE 3.
FT CARBOHYD 254 254 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 327 369 BY SIMILARITY.
FT LIPID 418 418 S-palmitoyl cysteine (By similarity).
FT LIPID 421 421 S-palmitoyl cysteine (By similarity).
SQ SEQUENCE 457 AA; 50871 MW; 57EDB344005A015 CRC64;

Query Match 68.1%; Score 1582.5; DB 1; Length 457;
Best Local Similarity 69.9%; Pred. No. 3.8e-101;
Matches 320; Conservative 34; Mismatches 65; Indels 39; Gaps 4;

QY 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGDVLTCTASQKSIQPHMKNSQIK 60
DB 1 MNGGIPFRHLLVLTQALLPAATGKTAVLGKGGVVELPCESTLKNVPPHMKTSQIK 60
QY 61 ILNGSGSFLTKGPSKLNDRADSRSLMDQGNFPLITIKLKTIEDSDTYICEVEDQKEEYQL 120
DB 61 ILGVNVTYTRGQSLTDRIDSKSSWDGSPFLIKARIEDSTYICEVESKKEEVEL 120
QY 121 LVFGLTANSDTHLLQGSLTTLTLESPGSSPSVQCRSPRKNIOGKTLVSQLELQDSG 180
DB 121 QVFGLTANPDTHLLQGSLTTLTLESPGSSPSVECTSPRKIRIRGKTLVSQLEIPDSG 180
QY 121 TWTCVLTQNKKEVEKIDIVLAFOKASSIVYKKEGEVSEFPLAFVEKLTGSGELMW 240
DB 181 TWKCTVFPHELELV-EEINIVLAFOQASSTVYKKEGEVSEFPLAFVEKLTGSGELCW 239
QY 241 QAERASSKSWITFDLKNKEVSVKRVTDQPKLQMGKKLPLHLTLPQALPOYAGSNTLTA 300
DB 240 QAERASSKSWITFDLKNKEVSVKRVTDQPKLQMGKKLPLHLTLPQALPOYAGSNTLTA 299
QY 301 LEAKTGKLGHOEVNVLVVRATQLOKNTLCEVWGPTSPKMLSLIKENKAKVSKREKPVW 360
DB 300 LKGGTGKLGHOEVNVLVVRATQLOKNTLCEVWGPTSPKMLSLIKENKAKVSKREKAVW 359
QY 361 LNPEAGMQCLSDSGOVLLESNIVLPTWSTPPVPRASALPAPTGSALDPOTASALP 420
DB 360 LNPEAGMQCLSDSGOVLLESNIVLPTWSTPPVPRASALPAPTGSALDPOTASALP 394
QY 421 DPPASALPALAVISFLGL----GLGV-ACVLARTR 453
DB 395 -----PMVLIVGVAGLAFGLGIFLCVRCRHR 424

RESULT 11
CD4_RABIT STANDARD; PRT; 459 AA.
AC P46630.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=92390370; PubMed=1518821;
RA Hage B.F., Sawadikosol S., Brown T.J., Lee K., Recker D.P.,
RA Kindt T.J.
RT "CD4 and its role in infection of rabbit cell lines by human
RT immunodeficiency virus type 1.";
```

```

RL Proc. Natl. Acad. Sci. U.S.A. 89:7963-7967(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M92840; AAA31198.1; -.
DR PIR; A46254; A46254.
DR HSP; P01730; IWR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TcANTIGEN.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS50835; IG-LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 459 T-CELL SURFACE GLYCOPROTEIN CD4.
FT DOMAIN 26 396 POTENTIAL.
FT TRANSMEM 397 419 POTENTIAL.
FT DOMAIN 420 459 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 129 IG-LIKE V-TYPE.
FT DOMAIN 130 208 IG-LIKE C2-TYPE 1.
FT DOMAIN 209 318 IG-LIKE C2-TYPE 2.
FT DOMAIN 319 374 IG-LIKE C2-TYPE 3.
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 41 113 POTENTIAL.
FT DISULFID 329 370 POTENTIAL.
FT LIPID 420 420 S-palmitoyl cysteine (By similarity).
FT LIPID 423 423 S-palmitoyl cysteine (By similarity).
SQ SEQUENCE 459 AA; 50886 MW; B32311C6D40013D CRC64;

Query Match 49.3%; Score 1147; DB 1; Length 459;
Best Local Similarity 58.0%; Pred. No. 2.5e-71;
Matches 240; Conservative 68; Mismatches 94; Indels 12; Gaps 6;

QY 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGDVLTCTASQKSIQPHMKNSQIK 60
DB 1 MNRRIYQCLLLVLPALLPAATGKTAVLGKGAIVELPCOSSQKNSVFNMGHAQVQK 60
QY 61 ILNGSG-----SFLTKGPSKLNDRADSRSLMDQGNFPLITIKLKTIEDSDTYICEVEDQKE 116
DB 61 ILNGQSSSSSFLTKGNSPLVNRVESKKNMMDQSPFLVTKLMDSDGYICEVGDKM 120
QY 117 EVOLVFGLTANSDTHLLQGSLTTLTLESPGSSPSVQCRSPRKNIOGKTLVSQLEL 176
DB 121 EVELLVFRLTANPTRLHLHGSLTLTLEGPVSGPSVQWMSPEKKIETGPTGMPRLRL 180
QY 177 QDSGTWCTCTV-LQNKQKVEFKIDIVLAFOKASSIVYKKEGEVSEFPLAFVEKLTGSG 235
DB 181 QDSGTWCSHLSPQDNKLELDIKIIVLGFPKASATVYKKEGEVSEFPLAFVEDESL--S 238
QY 236 GELMWQAERASSKSWITFDLKNKEVSVKRVTDQPKLQMGKKLPLHLTLPQALPOYAGSG 295
```

```

Db 239 GELMOWQVQASSAGSWVSFSLIEDRKVSQKILPDLKIQMSKGLPLSLTLPQALHRYAGSG 298
Qy 296 NLTLLALEAKTKLHOENVLVNMRATQLOKNTLCEVWGTSFKMLSLKLENKEAKVSKRE 355
Db 299 NLSLTLD--KGLTHQVSLVMLKVTVQVKNKLTCEVLGIDPKMLSLKLEDOEAKVS-TQ 355
Qy 356 KPVVYINDEAGWMOCLSDSGVLLSESNIKVLPWSTVPHPRASALPAPPTGSA 409
Db 356 KMWQVLDPKAKGTWQCLSSGQVLLSSKADVLATQLS--HQPTLLAGALGTA 407

RESULT 12
CD4_CANFA STANDARD; PRT; 463 AA.
AC P33705;
ID CD4_CANFA
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Leu-3).
GN CD4
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE OF 13-463 FROM N.A.
RC STRAIN=Beagle; TISSUE=Thymus;
RX MEDLINE=93192324; PubMed=7916632;
RA Milde K.F., Conner G.E., Minz D.H., Alejandro R.;
RT "Primary structure of the canine CD4 antigen.";
RL Biochim. Biophys. Acta 1172:315-318(1993).

RE SEQUENCE FROM N.A.
RC STRAIN=Beagle; TISSUE=Thymus;
RX MEDLINE=94378217; PubMed=8091416;
RA Gorman S.P., Frewin M.R., Cobbold S.P., Waldmann H.;
RT "Isolation and expression of cDNA encoding the canine CD4 and CD8
alpha antigens."
RL Tissue Antigens 43:184-188(1994).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed in macrophages and a subset of
CC T lymphocytes.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).

CC EMBL; L06130; AB02295.1; -
CC EMBL; X68565; -; NOT_ANNOTATED_CDS.
CC HSSP; P01730; 1MR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; ISS.
DR InterPro; IPR000973; CD4_TAG.
DR InterPro; IPR007110; Ig-Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 3.

```

```

DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1
FT CHAIN 25 463
FT DOMAIN 25 401 T-CELL SURFACE GLYCOPROTEIN CD4.
FT TRANSMEM 402 423 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 424 463 POTENTIAL.
FT DOMAIN 26 124 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 125 211 IG-LIKE C2-TYPE 1.
FT DOMAIN 212 321 IG-LIKE C2-TYPE 2.
FT DOMAIN 322 378 IG-LIKE C2-TYPE 3.
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 332 374 BY SIMILARITY.
FT LIPID 427 424 S-palmitoyl cysteine (by similarity).
FT LIPID 424 424 S-palmitoyl cysteine (by similarity).
FT CARBOHYD 123 123 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 389 389 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 463 AA; 51639 MW; 95805170CB4A833 CRC64;

Query Match 48.94; Score 1137; DB 1; Length 463;
Best Local Similarity 52.54; Pred. No. 1.2e-70;
Matches 242; Conservative 67; Mismatches 100; Indels 52; Gaps 8;

Qy 1 MNRGVPRLHLVLVQLALPAAATQGNKVVVGKKGDTVELTCTASQKKSIGPHMKNNSQIK 60
Db 1 MNRGEARHLLMLQVLMLPAAVTPVREVLVGRKAGDAVELPCQSQKKNHFNRRDSSMWQ 60
Qy 61 ILNQGSEFLTKGSKLNDRAISRSLMDQGNFLLIKNTKLESDITYICEVEQKEVQL 120
Db 61 ILNQGSEFWTVGSSRLKRVESKKNLMDQGFPLVKLEVAQSGIYFCPT-DKROEVEL 119
Qy 121 LVFGLTP-----NSDTHLQGGSLTLETSPGSSSPVOCRSPRGNIQGGKTLAV 171
Db 120 LVFNLTAKMDSSGSSSSNRLQGGQLTTLTENPSSSSVWKGPGNKSRRGGQNL 179
Qy 172 SOLELDPSGTWCTVLQONKKEFKIDIVLAFOKASIVYKKEGQVSEFPLAFTVEK 231
Db 180 SWPELDQGTWCTIISGQKTEFNFNLVLAQKSNFTYARAGDQVSEFPLSFEDEN 239
Qy 232 LTSGSELMWAQEAASSKSWITFDLKNKEVSVRVTDPKLQMGKPLHLTPQALPQY 291
Db 240 LV--GELRWQAQAGASSSLMWISFTLENRKLMSKEAHPKLQMKESLPRLFTLPQVLSRY 297
Qy 292 AGSGNLTALAEATGKLBHOENVLVNMRATQLOKNTLCEVWGTSFKMLSLKLENKEAKV 351
Db 298 AGSGILTLNL-AK-GTLYQENVLVNMRANSSQNNLTCEVLCPTSPETLLSLNKEQAKV 355
Qy 352 SKKEKPVVYINDEAGWMOCLSDSGVLLSESNIKVLPWSTVPHPRASALPAPPTGSA 411
Db 356 SKQQKLVVWVDPGEGTWQCLSDKXKVLASSLNV-----SSPV----- 394
Qy 412 DPQTASALPDPAPASALPALA-----VISFLGLGLGVNC 447
Db 395 -----VIKSPKFLATITLGGILGLLILIGLCVFC 423

RESULT 13
CD4_RAT STANDARD; PRT; 457 AA.
ID CD4_RAT
AC P05540;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Leu-3) (W3/25 antigen).
GN CD4.

```


OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=87175535; PubMed=3104900;
RA Clark S.J., Jeffries W.A., Barclay A.N., Gagnon J., Williams A.F.,
RT "Peptide and nucleotide sequences of rat CD4 (W3/25) antigen:
RT evidence for derivation from a structure with four
RT immunoglobulin-related domains.";
RT Proc. Natl. Acad. Sci. U.S.A. 84:1649-1653(1987).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 210-399.
RA Brady R.L., Dodson E.J., Dodson G.G., Lange G., Davis S.J.,
RA Williams A.F., Barclay A.N.;
RT "Crystal structure of domains 3 and 4 of rat CD4: relation to the
RT NH2-terminal domain.";
RT Science 260:979-983(1993).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL: M15768; AAA0901.1; -
DR PIR: A27449; A27449.
DR PDB: 1CID; 15-JUL-93.
DR GlycoSiteDB; P05540; -
DR GO: GO:0042101; C1T-cell receptor complex; ISS.
DR GO: GO:0015026; F1:receptor activity; ISS.
DR GO: GO:0042289; F1:MHC class II protein binding; ISS.
DR GO: GO:0006955; P1:immune response; ISS.
DR GO: GO:0045086; P1:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO: GO:0032121; P1:T-cell differentiation; ISS.
DR GO: GO:0045058; P1:T-cell selection; ISS.
DR GO: GO:0007169; P1:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro: IPR000973; CD4 TCAG.
DR InterPro: IPR007110; Ig-Like.
DR InterPro: IPR003599; Ig.
DR Pfam: PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00409; IG_2.
DR PROSITE; PS00835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure.
FT SIGNAL 1 27
FT CHAIN 28 457
FT DOMAIN 28 394
FT TRANSMEM 395 417
FT DOMAIN 418 457
FT DOMAIN 28 127
FT DOMAIN 128 206
FT DOMAIN 207 316
FT DOMAIN 317 374
FT CARBOHYD 186 186
FT CARBOHYD 297 297
FT CARBOHYD 392 392
FT DISULFID 43 111
FT DISULFID 158 187
FT DISULFID 370 370
FT LIPID 418 421
FT LIPID 421 421
T-CELL SURFACE GLYCOPROTEIN CD4.
EXTRACELLULAR (POTENTIAL).
POTENTIAL.
CYTOPLASMIC (POTENTIAL).
IG-LIKE V-TYPE.
IG-LIKE C2-TYPE 1.
IG-LIKE C2-TYPE 2.
IG-LIKE C2-TYPE 3.
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
S-palmitoyl cysteine (By similarity).
S-palmitoyl cysteine (By similarity).

FT STRAND 213 217
FT TURN 218 219
FT STRAND 222 225
FT STRAND 225 243
FT STRAND 235 243
FT TURN 252 258
FT STRAND 259 260
FT STRAND 261 265
FT STRAND 274 275
FT STRAND 278 278
FT TURN 279 280
FT STRAND 282 285
FT STRAND 290 292
FT HELIX 294 301
FT STRAND 306 319
FT STRAND 325 331
FT STRAND 338 344
FT TURN 345 346
FT STRAND 349 353
FT STRAND 357 361
FT STRAND 367 374
FT TURN 375 376
FT STRAND 377 385
SQ SEQUENCE 457 AA; 51437 MW; 477BE157D30954C1 CRC64;
Query Match 43.0%; Score 999; DB 1; Length 457;
Best Local Similarity 52.9%; Pred. No. 3.4e-61;
Matches 207; Conservative 61; Mismatches 115; Indels 8; Gaps 5;
QY 1 MNRGVPPRLH--LLVLQALPAATQGNKVVAGKGDVTELCTASQKRSIQFHWKNSQ 58
DB 1 MCRGSPFRLHLLPILLQSLKLVLTQKTVVLGKGGSALEPCSTSRSSAFWKSDDQ 60
QY 59 IKIIGNQSFTKPSKLNDAISRSLMDQGNPILLIKLIKIDSDTIYIEVDQKEV 118
DB 61 KTIIGYKXKLIKLSLEYSRFRDSRKAMERGSPLIINKLRMBDSQTYVELNKKEV 120
QY 119 QLVFGLTANSDDTHLQGSITLTLES--PGSSPSVOCRSRPGKNIOGKTLVSQLEQ 177
DB 121 ELWFRVTFNPGTLLQGSITLTLDNSPKVSDPIECRKHSSIVDSKAFSTHSRLIQ 180
QY 178 DSGTWTCTVLONQKKEFKIDIVVAFQKASSIYKKEGQVEFSFPLAFTVEKLTSGE 237
DB 181 DSGIINCTVLTNQKHSPDKLVLGFASTSITAYKSEGEAEFSPLNLGEESL--QGE 238
QY 238 LMQAERASSKSSKITTPLDKRKESYKRVYODPFLQNGKPLHLPLPOLPQVAGSNTL 297
DB 239 LRWAEKAPSSQSWITFSLKNQKVSVOKSTSNPKFQSETLPLTLQIPQVSLQFAGSNTL 298
QY 298 TLAEAKTGKLGHOENVLVVWRATOLQKN-LTCEVWGPSPKLMLSKLENKEAKVSKREK 356
DB 299 TLITLD--RGILYQENVLVWVKVTOPDSNTLTCEVWGPSPKMRLLKQENQEARVSKROK 356
QY 357 PFWVLNPEAGMWCCLISDSGQVLESNIKVL 387
DB 357 VIQVAPAGVWQCCLISGGEVKNDSKIQVL 387
RESULT 14
CD4_MOUSE
ID CD4_MOUSE STANDARD; PRT; 457 AA.
AC P06332;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3) (T-cell differentiation antigen L3T4).
GN CD4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=87018645; PubMed=3094146;
 RA Tourville B., Gorman S.D., Field E.H., Hunkapiller T., Parnes J.R.;
 RT "Isolation and sequence of L3T4 complementary DNA clones: expression
 in T cells and brain.";
 RL Science 234:610-614(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87115821; PubMed=3027575;
 RA Littman D.R., Gettner S.N.;
 RT "Unusual intron in the immunoglobulin domain of the newly isolated
 RT murine CD4 (L3T4) gene.";
 RL Nature 325:453-455(1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88152875; PubMed=3326818;
 RA Parnes J.R., Hunkapiller T.;
 RT "L3T4 and the immunoglobulin gene superfamily: new relationships
 RL between the immune system and the nervous system.";
 RL Immunol. Rev. 100:109-127(1987).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Brain;
 MEDLINE=88041159; PubMed=2823269;
 RA Gorman S.D., Tourville B., Parnes J.R.;
 RT "Structure of the mouse gene encoding CD4 and an unusual transcript
 RT in brain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648(1987).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98112780; PubMed=9445485;
 RA Anselmi-Lari M.A., Oelsten J.C., Schwartz S., Zhang Z., Muzny D.M.,
 RA Lu J., Gorrell J.H., Chnault A.C., Belmont J.W., Miller W.,
 RA Gibbs R.A.;
 RT "Comparative sequence analysis of a gene-rich cluster at human
 RT chromosome 12p13 and its syntenic region in mouse chromosome 6.";
 RL Genome Res. 8:29-40(1998).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX STRAIN=C57BL/6J; TISSUE=Mammary gland;
 MEDLINE=22386257; PubMed=12477932;
 RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeebfer B., Buelow K.H., Scheffer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stepelson M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein W.J., Usdin T.B., Toshiyuki S., Carninci P., Franke C.,
 RA Rana S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kertman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield V.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [7]
 RP SEQUENCE OF 27-43.
 RX MEDLINE=8616694; PubMed=3082751;
 RA Classon B.J., Tsagaratos J., Kirschbaum L., Maddox J., McKay C.R.,
 RA Brandon M., McKenzie I.F.C., Walker I.D.;
 RT "The L3T4 antigen in mouse and the sheep equivalent are
 RT immunoglobulin-like.";
 RL Immunogenetics 23:129-132(1986).
 RN [8]
 RP DISULFIDE BONDS.
 RX MEDLINE=86233454; PubMed=3086886;
 RA Classon B.J., Tsagaratos J., McKenzie I.F.C., Walker I.D.;
 RT "The L3T4 antigen in mouse and the sheep equivalent are
 RT immunoglobulin-like.";
 RL Immunogenetics 23:129-132(1986).

RT "Partial primary structure of the T4 antigens of mouse and sheep:
 RT assignment of intrachain disulfide bonds.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4499-4503(1986).
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
 CC receptor interaction. May regulate T-cell activation.
 CC -1- SUBUNIT: Associates with p56-lck (by similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=P06332-1; Sequence=Displayed;
 CC Name=2; Synonyms=Brain-specific;
 CC IsoId=P06332-2; Sequence=VSP_002489;
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL: M36850; AAA39401.1; -;
 CC EMBL: M13816; AAA37267.1; -;
 CC EMBL: X04836; CAA28539.1; -;
 CC EMBL: M36851; AAA39402.1; -;
 CC EMBL: M17080; AAA37403.1; -;
 CC EMBL: M17078; AAA37403.1; JOINED.
 CC EMBL: M17079; AAA37403.1; JOINED.
 CC EMBL: AC002397; AAC36010.1; -;
 CC EMBL: BC039137; AAC39137.1; -;
 CC PIR: A02110; RWMST4.
 CC HSPSP; P01730; 1MBR.
 CC MGD; MGI:88335; Cd4.
 CC GO: GO:0042101; C:T-cell receptor complex; ISS.
 CC GO: GO:0015026; F:coreceptor activity; ISS.
 CC GO: GO:0042289; F:MHC class II protein binding; ISS.
 CC GO: GO:0006955; P:immune response; ISS.
 CC GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
 CC GO: GO:0030217; P:T-cell differentiation; ISS.
 CC GO: GO:0045058; P:T-cell selection; ISS.
 CC GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
 CC InterPro: IPR000973; CD4 TCRg.
 CC InterPro: IPR007110; Ig-Like.
 CC InterPro: IPR003596; Ig_V.
 CC Pfam; PF00047; Ig_2.
 CC PRINTS; PR00692; CD4TCANTIGEN.
 CC SMART; SM00406; IGV_1.
 CC PROSITE; PS50835; IG_LIKE_1.
 CC Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
 CC Immune response; Repeat; Signal; Lipoprotein; Palmitate;
 CC Alternative splicing.
 CC KW
 CC CHAIN 1
 CC SIGNAL 26
 CC DOMAIN 27 457
 CC TRANSMEM 27 394
 CC DOMAIN 417
 CC TRANSMEM 395 417
 CC DOMAIN 418 457
 CC DOMAIN 27 128
 CC DOMAIN 129 207
 CC DOMAIN 208 317
 CC DOMAIN 318 374
 CC CARBOHYD 187 187
 CC CARBOHYD 298 298
 CC CARBOHYD 323 323
 CC CARBOHYD 392 392
 CC CARBOHYD 42 112
 CC DISULFID 159 188
 CC DISULFID 328 370
 CC LIPID 418 418
 CC LIPID 418 418
 CC LIPID 421 421
 CC VARSPLIC 1 240
 CC
 CC T-CELL SURFACE GLYCOPROTEIN CD4.
 CC EXTRACELLULAR (POTENTIAL).
 CC POTENTIAL.
 CC CYTOPLASMIC (POTENTIAL).
 CC IG-LIKE V-TYPE.
 CC IG-LIKE C2-TYPE 1.
 CC IG-LIKE C2-TYPE 2.
 CC IG-LIKE C2-TYPE 3.
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC S-palmitoyl cysteine (By similarity).
 CC S-palmitoyl cysteine (By similarity).
 CC Missing (in isoform 2).

```

FT SEQUENCE 457 AA; 51296 MW; 181DA7527CB00F33 CRC64; /FTID=VSP 002489.
SQ Query Match 42.7%; Score 993; DB 1; Length 457;
Beet Local Similarity 53.8%; Pred. No. 8,8e-61;
Matches 211; Conservative 62; Mismatches 109; Indels 10; Gaps 7;

QY 1 MNRGVPFRF-LTLVQLALPPAATGKNKVLKKGDDTVLTCTASQOKSIOPHMNSNOI 59
DB 1 MRAISLRLLQLLQSLQALVQKTLVLGSGESABLPSSQOKITVFTWKFSDR 60
QY 60 KILGNQGS-SFLTKG--PSKLNDRADSRSLMDQGNFPLIKNLKIEDSPVYICEVEDQKE 116
DB 61 KILGKHGKGVLRGSGPSQF-DRPQSKGAWKSGPPLINKLKNMEDSGTYICELENKE 119
QY 117 EQVLVFLGLTANSDPHLLQSGSLTLTLES-PPGSSPSVQCSRPKNIOGKTLVSQLE 175
DB 120 EVELWVFVKYTFSPGTSLLQSGSLTLTLDNSKVSNDPLTECKKHKGVSGSKVLSMSNLR 179
QY 176 LQDSGWTCTVQLONOKKVEFKIDIVLAFQKASIVYKKGQVESPFLAFTVEKLGS 235
DB 180 VQDSDFMNCVTVLDDQKMFGLTSLVLFQSTAIRAYSGESABSPFLNFAER--NGW 237
QY 236 GELMWQAERASSSSKSNITFDLNKKEVSVKVTQDPKLGKGLPLHLTLPLQALPYAGSG 295
DB 238 GELMWKAEDSFQFPQWISFSINKKEVSGKTKDLQKLTPLTLIKIPQVSLQFASG 297
QY 296 NITLLEAKTGKLVHQRNLVVRAPQLOKNTLCEVWGPSPKMLSLKLENKAVSKRE 355
DB 298 NITLTLTD--KGTLDHEVNLVWVKVQALNNTLTCEVWGPSPKMLTLKQENQEARVSEBQ 355
QY 356 KPVWVLNPEAGMWQCLSDSGVLLSESNIKVL 387
DB 356 KVVQVAPETGLMQCLSEGDKVKNDSRIQVL 387

RESULT 15
CD7_HUMAN STANDARD; PRT; 240 AA.
AC P09564;
DT 01-MAR-1989 (rel. 10, Created)
DT 01-MAR-1989 (rel. 10, Last sequence update)
DT 10-OCT-2003 (rel. 42, Last annotation update)
DE T-cell antigen CD7 precursor (GP40) (T-cell leukemia antigen) (TP41)
DE (Leu-9).
GN CD7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN 1[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8811517; PubMed=3501369;
RA Aruffo A., Seed B.;
RT "Molecular cloning of two CD7 (T-cell leukemia antigen) cDNAs by a
RT COS cell expression system.";
RN 2[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9110576; PubMed=1703303;
RA Schanberg L.E., Fleener D.E., Kurtzberg J., Haynes B.F., Kaufman R.E.;
RT "Isolation and characterization of the genomic human CD7 gene:
RT structural similarity with the murine Thy-1 gene.";
RN 3[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22388257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Schenfer C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buecok K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heisch F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

```

```

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedon T.B., Toshiyuki S., Carrini P., Prange C.,
RA Raba S.S., Loguallano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gamarate P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RN 4[4]
RP SEQUENCE OF 205-240 FROM N.A.
RX MEDLINE=91267564; PubMed=1711009;
RA Yoshikawa K., Seto M., Ueda R., Obata Y., Notake K., Yokochi T.,
RA Takahashi T.;
RT "Molecular cloning of the gene coding for the human T cell
RT differentiation antigen CD7.";
RN 5[5]
RP Immunogenetics 33:352-360(1991).
TOPOLOGY.
RX MEDLINE=90063052; PubMed=2479685;
RA Ware R.E., Searce R.M., Dietz M.A., Starmer C.F., Parker T.J.,
RA Haynes B.F.;
RT "Characterization of the surface topography and putative tertiary
RT structure of the human CD7 molecule.";
RN 6[6]
RP J. Immunol. 143:3632-3640(1989).
INTERACTION WITH SECTM1.
RX MEDLINE=20119303; PubMed=10652336;
RA Lyman S.D., Escobar S., Rousseau A.-M., Armstrong A., Fanslow W.C.;
RT "Identification of CD7 as a cognate of the human K12 (SECTM1)
RT protein.";
RN 7[7]
RP J. Biol. Chem. 275:3431-3437(2000).
CC 1- FUNCTION: Not yet known.
CC 1- SUBUNIT: Interacts with SECTM1.
CC 1- SUBCELLULAR LOCATION: Type I membrane protein.
CC 1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC 1- DATABASE: NAME=PROW; NOTE=CD guide CD7 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd7.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06180; CAA29546.1; -
DR EMBL; BC37271; AAA51953.1; -
DR EMBL; BC009293; AAA09293.1; -
DR EMBL; BC013297; AAA13297.1; -
DR EMBL; D00749; BAA00646.1; -
DR EMBL; D00749; BAA00646.1; JOINED.
DR EMBL; D00748; BAA00646.1; JOINED.
DR PIR; A39016; A39016.
DR HSRP; P01607; IREI.
DR Genew; HGNC:1695; CD7.
DR MIM; 186820; -
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0005866; C:plasma membrane; TAS.
DR GO; GO:0004872; F:receptor activity; TAS.
DR GO; GO:0006955; P:immune response; TAS.
DR GO; GO:0042110; P:T-cell activation; TAS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. .; NNS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003599; IG.

```

```

DR Pfam; PF00047; Ig; 1.
DR SMART; SM00409; IG; 1.
DR PROSITE; PS50835; IG LIKE; 1.
KW T-cell; Signal; Immune response; Antigen; Transmembrane; Glycoprotein;
KW Immunoglobulin domain; Receptor; Lipoprotein; Palmitate; Repeat.
FT SIGNAL 1 25
FT CHAIN 26 240
FT DOMAIN 26 180 T-CELL ANTIGEN CD7.
FT TRANSMEM 181 201 EXTRACELLULAR (PROBABLE).
FT DOMAIN 202 240 PROBABLE.
FT DOMAIN 226 240 CYTOPLASMIC (PROBABLE).
FT DOMAIN 26 130 IG-LIKE.
FT DOMAIN 145 180 4 X 9 AA TANDEM REPEATS, POTENTIAL SPACER
FUNCTION.
FT DISULFID 35 142 POTENTIAL.
FT DISULFID 48 114 POTENTIAL.
FT LIPID 198 198 S-palmitoyl cysteine.
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .).
FT REPEAT 145 153 1.
FT REPEAT 154 162 2.
FT REPEAT 163 171 3.
FT REPEAT 172 180 4.
SQ SEQUENCE 240 AA; 25409 MW; EBBCE08279552108 CRC64;

Query Match 12.6%; Score 293; DB 1; Length 240;
Best Local Similarity 40.3%; Pred. No. 3.3e-13;
Matches 91; Conservative 18; Mismatches 49; Indels 68; Gaps 9;

QY 270 PRLQWKKLPLHLPLQALPOVAGSGNLTALAEATGKLGHOENVVMARATQLOKNLTCE 329
DB 5 PRLTL---LPLLLALARGPG-----ALAAQEVQOSPCHCTTPVGAS---VNITCS 49
QY 330 VWGPTSPKLM.LSLKENKAVSKREKPVVVLNPEAG----- 366
DB 50 TSGGARGIYLNQL-----GPOQDIIVYEDGVPTDRFRGRIDPSSQDMLT 98
QY 367 --MMOCLISDSG---QVLESNT-----KVLPT-----W-STVHPRASALPAPPTG 407
DB 99 ITMRLQLSDTGTVCQAITEVNVSGTLVLYTEEOGQMRHRCSDADPPRASALPAPPTG 158

QY 408 SALPDPOTASALPDPASALPALAVISFLIGLGVACVLAARR 453
DB 159 SALPDPOTASALPDPASALPALAVISFLIGLGVACVLAARTQ 204

RESULT 16
VCAL RAT STANDARD; PRT; 739 AA.
AC P29534;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vascular cell adhesion protein 1 precursor (V-CAM 1).
GN VCAM1 OR VCAM-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA MEDLINE=92181437; PubMed=1371918;
RA Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wyke M.,
RA Burley L., Miyake K., Kincaid P., Lobb R.;
RT "Cloning of murine and rat vascular cell adhesion molecule-1.";
RL Biochem. Biophys. Res. Commun. 183:163-169(1992).
CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
CC INTEGRIN VLA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A
CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
CC EMIGRATION TO SITES OF INFLAMMATION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

```

```

CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as
CC well as on macrophage-like and dendritic cell types in both normal
CC and inflamed tissue.
CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; M84488; AAA42332.1; -.
CC PIR; J50675; J50675.
CC HSSP; P19320; IYCA.
CC InterPro; IPR003987; ICAM VCAM-1.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003598; Ig_c2.
CC InterPro; IPR003989; VCAM-1.
CC Pfam; PF00047; Ig; 5.
CC PRINTS; PR01472; ICAMVCAM1.
CC PRINTS; PR01474; VCAM1.
CC SMART; SM00408; IGc2; 3.
CC PROSITE; PS50835; IG LIKE; 5.
CC KW Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
CC Repeat; Signal.
FT SIGNAL 1 24 PROBABLE.
FT CHAIN 25 739 VASCULAR CELL ADHESION PROTEIN 1.
FT DOMAIN 25 698 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 699 720 POTENTIAL.
FT DOMAIN 721 739 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 119 212 IG-LIKE C2-TYPE 1.
FT DOMAIN 223 309 IG-LIKE C2-TYPE 2.
FT DOMAIN 312 397 IG-LIKE C2-TYPE 3.
FT DOMAIN 408 506 IG-LIKE C2-TYPE 4.
FT DOMAIN 514 595 IG-LIKE C2-TYPE 5.
FT DOMAIN 601 682 IG-LIKE C2-TYPE 6.
FT DISULFID 47 95 BY SIMILARITY.
FT DISULFID 52 99 BY SIMILARITY.
FT DISULFID 137 195 BY SIMILARITY.
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 739 AA; 81246 MW; 5C6085A1A1B100C CRC64;

Query Match 7.0%; Score 162; DB 1; Length 739;
Best Local Similarity 22.0%; Pred. No. 0.0013;
Matches 86; Conservative 60; Mismatches 155; Indels 90; Gaps 14;

QY 33 KEDVELITCTASQKKSIOFWKNSNQKILGNGSFLTPKPSKLNDRADRSRLMDGNF 92
DB 238 EGAAVTTCASEGLPAPEIFWSKK-----LDNGVLDTL-----SGNA 274
QY 93 PLIKIKIKIDSDTYICE-----VEDQKEVQLV-----FGTANSDTLIGOSLTLT 142
DB 275 TLTILAKMEDSGIYCEGNLVGRDKTEVELLVQEKRFVVDISPSQVAAQVGDSDVLT 334
QY 143 LESPSSPSVQCRSPKKNIOG-----GKTLVSQLELDOSGTCTVLONQKVEF 195
DB 335 CAAVAGCSPSPFMSRQTQDPSLNGSVNDEGATSTLTSPVGEVBHSLVLTGQGRRLTEK 394
QY 196 KIDIVLAFOKASIVYKKEGEOVRSPLA-----FTVEKLTGSG 236
DB 395 TIQEVYVSF-----PDPEIETISGPLVHGRPVTVNCTVNVVYFPDHLIELLNGET 445
QY 237 ELWMOQARRAS-SKSWITFLPKNKEVSVKRVTDPPKLGWKKL-----PLHLTLPOALP-Q 290
DB 446 TLANKFLREISGTGS---LETKSLNMTFTPAED---TGKALVCLAKLHSSQWSESPKQ 498

```

```

OY 291 VAGSGLTALBAKTKLHQEVNLVWMRATQKQKLTCEWQPTSPKMLSTKLENKAK 350
DB 499 ROSTQTLVYNVAPKPTTIVSPSPVPEESPV--NLTCSDGPFPPKILMSQNLKGBLQ 556
OY 351 VSKREKPVWVNLPEAGMOCILSDSGQVLE 381
DB 557 PLSQ-----NTLSFMATKEDSGIYVCE 580.

RESULT 17
VCAM1 HUMAN STANDARD; PRT; 739 AA.
ID P19320;
AC 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vascular cell adhesion protein 1 precursor (V-CAM 1) (CD106 antigen)
DE (INCAM-100).
GN VCAM1 OR L1CAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=umbilical vein;
RX MEDLINE=91016951; PubMed=1699207;
RA Pote T., Newman W., Gopal T.V.;
RT "Full length vascular cell adhesion molecule 1 (VCAM-1).";
RL Nucleic Acids Res. 18:5901-5901 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90090619; PubMed=2688898;
RA Osborn L., Hesston C., Tizard R., Vassallo C., Luhnowsky S.,
RA Chi-Rosco G., Lobb R.;
RT "Direct expression cloning of vascular cell adhesion molecule 1, a
RT cytokine-induced endothelial protein that binds to lymphocytes.";
RL Cell 59:1203-1211 (1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=9152090; PubMed=1715583;
RA Cybulsky M.I., Fries J.W.U., Williams A.J., Sultan P., Eddy R.,
RA Byers M., Shows T., Gimbrone M.A. Jr., Collins T.;
RT "Gene structure, chromosomal location, and basis for alternative mRNA
RT splicing of the human VCAM1 gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:7859-7863 (1991).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91201302; PubMed=1707873;
RA Hesston C., Tizard R., Vassallo C., Schiffer S.B., Goff D., Moy P.,
RA Chi-Rosco G., Luhnowsky S., Lobb R., Osborn L.;
RT "Cloning of an alternate form of vascular cell adhesion molecule-1
RT (VCAM1).";
RL J. Biol. Chem. 266:6682-6685 (1991).
RN [5]
RP SEQUENCE FROM N.A., AND VARIANTS PHE-318; ALA-384; ALA-413 AND
RP LEU-716.
RA Rieder M.J., Armet T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,
RA Rajkumar N.R., Toch E.J., Yi Q., Nickerson D.A.;
RT Submitted (Aug-2002) to the EMBL/Genbank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM LONG).
RC TISSUE=Retinal pigment epithelium;
RX MEDLINE=2238857; PubMed=12477932;
RA Struhsberg R.L., Feingold E.A., Grose L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopfins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Pange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abriemson R.D., Mullighy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

```

```

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kottman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhe D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 25-226.
RX MEDLINE=95147978; PubMed=7531291;
RA Jones E.Y., Harlos K., Bottomley M.J., Robinson R.C., Driscoll P.C.,
RA Edwards R.W., Clements J.M., Dudgeon I.J., Stuart D.I.;
RT "Crystal structure of an integrin-binding fragment of vascular cell
RT adhesion molecule-1 at 1.8-A resolution.";
RL Nature 373:539-544 (1995).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
RX MEDLINE=95296382; PubMed=7539925;
RA Wang J.-H., Pepinsky R.B., Stehle T., Liu J.-H., Karpusas M.,
RA Browning B., Osborn L.;
RT "The crystal structure of an N-terminal two-domain fragment of
RT vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on
RT the domain 1 C-D loop can inhibit VCAM-1-alpha 4 integrin
RT interaction.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:5714-5718 (1995).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
RA Wang J.-H., Stehle T., Pepinsky R.B., Liu J.-H., Karpusas M.,
RA Osborn L.;
RT "Structure of a functional fragment of VCAM-1 refined at 1.9-A
RT resolution.";
RL Acta Crystallogr. D 52:369-379 (1996).
CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
CC INTEGRIN VLAM ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A
CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
CC EMIGRATION TO SITES OF INFLAMMATION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=Long;
CC IsoId=P19320-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P19320-2; Sequence=VSP_002580;
CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as
CC well as on macrophage-like and dendritic cell types in both normal
CC and inflamed tissue.
CC -1- INDUCTION: By cytokines (e.g. IL-1, TNF-alpha).
CC -1- PTM: Sialoglycoprotein.
CC -1- DISEASE: May play an important role in the genesis of
CC artherosclerosis and rheumatoid arthritis.
CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD106 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd106.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; X53051; CAA37218.1; -
CC EMBL; M30257; AAA51917.1; ALT TERM.
CC EMBL; M73355; AAA61270.1; -
CC EMBL; M60335; AAA61269.1; -

```

```

DR EMBL; AF536818; AAM96190.1; -
DR EMBL; BC017276; AAH17276.1; -
DR PIR; A41288; A41288.
DR PIR; B41288; B41288.
DR PDB; 1VCA; 15-SEP-95.
DR PDB; 1VSC; 20-JUN-96.
DR PDB; 1IJ9; 07-NOV-01.
DR Genew; HGNC:12663; VCAM1.
DR MIM; 192225; -.
DR InterPro; IPR003987; ICAM VCAM-1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003989; VCAM-1.
DR Pfam; PF00047; Ig; 6.
DR PRINTS; PR01472; ICAMVCAM1.
DR PRINTS; PR01474; VCAM1.
DR SMART; SMO0408; IGC2; 3.
DR PROSITE; PS50835; IG_LIKE; 5.
KW Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
Repeat; Signal; Alternative splicing; Polymorphism; 3D-structure.
FT SIGNAL 1 24
FT CHAIN 1 25 739
FT DOMAIN 25 698
FT TRANSMEM 699 720
FT DOMAIN 721 739
FT DOMAIN 25 105
FT DOMAIN 109 212
FT DOMAIN 223 309
FT DOMAIN 312 399
FT DOMAIN 408 506
FT DOMAIN 511 595
FT DOMAIN 600 684
FT DISULFID 47 95
FT DISULFID 52 99
FT DISULFID 137 195
FT CARBOHYD 273 273
FT CARBOHYD 365 365
FT CARBOHYD 417 417
FT CARBOHYD 463 463
FT CARBOHYD 531 531
FT CARBOHYD 561 561
FT VAAPPLIC 310 402
FT VARIANT 318 318
FT VARIANT 384 384
FT VARIANT 413 413
FT VARIANT 716 716
FT STRAND 26 30
FT STRAND 34 38
FT TURN 39 40
FT STRAND 43 50
FT STRAND 56 61
FT TURN 62 63
FT STRAND 70 74
FT TURN 75 76
FT STRAND 77 82
FT HELIX 87 89
FT TURN 92 99
FT STRAND 100 101
FT STRAND 102 114
FT STRAND 120 123
Query Match 6.7%; Score 155; DB 1; Length 739;
Best Local Similarity 21.9%; Pred. No. 0.004;
Matches 91; Conservative 71; Mismatches 166; Indels 88; Gaps 18;
8 RHLLVLTALLLPAT--QGKRVVIGKGGDYVELCTASQKKSIOFHMKNSNQIKILNQ 65
Db ROAVKELOVVISPKNTVISVNPSTKLQGGSVTMTCSSEGLPAPERFW----- 258

```

```

QY 66 GSFLTKGPSKLNRADSRRLSDQGNFLLIRKLIKEDSDPTLYCE---VEDQKEEYQL 121
D 259 -----SKKLNGNLQH-----SGNATLTLIMBRMDSGICYCEGNLIGKRRKEVELI 307
QY 122 V----FGLTANSTHLLIQ--GSLTLTLESPPSSPSVQCRSPKNGIQ-----GKT 168
D 308 VQEKPFVVEISPPRRIAQIGDSVMTLCSVMGCESSPSFRRTOIDSPLSKRVSEGTNST 367
QY 169 LSVSQLELDQSGTWTCTVLQNOKKVEFKIDIVLAPQKASIIYK---KEGEQVEFS--- 222
D 368 LTLSPVSEFNEHSYLLCTVTGHHKLEKIGVLEYSPRDEIEMSGGLVNGSSVYTSCKV 427
QY 223 ---FPL-AFTVEKLTSGELMQAERASSKSWITF---DLNKEVSVRVYTDPPQLQ- 273
D 428 PSVYPLDRLEIELKGETTL-----ENIEFLBDTDMKSLKSLMEFTPIETD 476
QY 274 MGKTL---PLHLTLPOALFOYAGSGNLTALAEAKTGKLEHENVLVYMRATQLOK---N 325
D 477 TGRALVQQAHLHTDMEFEFKQROF---TOTLVYVNAF--RDTTVLVSPSIIIEGSSVN 531
QY 326 LTCEWVGPSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMMOCLLSDSGVYLE 381
D 532 WTCLSGCFAPAKTLWSRQLPNGLQPLSENATLTLSTK-----MEDSGVYLCE 580

RESULT 18
NCAM2 HUMAN STANDARD; PRT; 761 AA.
ID NCAM2 HUMAN
AC P13592; P13593;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 1, 120 kDa isoform precursor (N-CAM 120)
DE (NCAM-120) (CD56 antigen).
GN NCAM1 OR NCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM N-CAM 120).
RC TISSUE=Skeletal muscle;
RX MEDLINE=89305258; PubMed=3253057;
RA Barton C.H., Dickson G., Gower H.J., Rowett L.H., Putt W.,
RA Elsom V., Moore S.E., Goridis C., Walsh F.S.;
RT "Complete sequence and in vitro expression of a tissue-specific
RT phosphatidylinositol-linked N-CAM isoform from skeletal muscle.";
RL Development 104:165-173(1988).
RN [2]
RP SEQUENCE OF 491-761 FROM N.A. (ISOFORM N-CAM 120).
RC TISSUE=Skeletal muscle;
RX MEDLINE=87301755; PubMed=2887295;
RA Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
RA Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;
RT "Human muscle neural cell adhesion molecule (N-CAM): identification
RT of a muscle-specific sequence in the extracellular domain.";
RL Cell 50:1119-1130(1987).
RN [3]
RP SEQUENCE OF 491-655 FROM N.A. (ISOFORM C).
RX MEDLINE=89077552; PubMed=3203385;
RA Gower H.J., Barton C.H., Elsom V.L., Thompson J., Moore S.E.,
RA Dickson G., Walsh F.S.;
RT "Alternative splicing generates a secreted form of N-CAM in muscle
RT and brain.";
RL Cell 55:955-964(1988).
CC -1- FUNCTION: This protein is a cell adhesion molecule involved in
CC neuron-neuron adhesion, neurite fasciculation, outgrowth of
CC neurites, etc.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC Event=Alternative splicing; Named isoforms=3;
CC Name=N-CAM 120;

```

```
CC      IsoId=P13592-2; Sequence=Displayed;
CC      Name=N-CAM 140;
CC      IsoId=P13591-1; Sequence=External;
CC      Name=C; Synonyms=Secreted;
CC      IsoId=P13592-1; Sequence=VSP_002587;
CC      -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC      -1- SIMILARITY: Contains 2 fibronectin type III domains.
CC      -1- DATABASE: NAME=PROW; NOTE=CD guide CD56 entry;
CC      WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd56.htm".
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL, X16841; CAA34739.1; -
CC      DR      EMBL, M17409; AAA59912.1; -
CC      DR      EMBL, M22094; AAA59910.1; -
CC      DR      EMBL, M22092; AAA59911.1; -
CC      DR      EMBL, M22091; AAA59911.1; JOINED.
CC      DR      PIR, A31635; A31635.
CC      DR      PIR, S07784; JHUNG.
CC      DR      Genew; HGNC:7656; NCAM1.
CC      DR      MIM: 116930; -
CC      DR      GO; GO:0016021; C: integral to membrane; TAS.
CC      DR      GO; GO:0005886; C: plasma membrane; TAS.
CC      DR      InterPro; IPR008957; FN_III-like.
CC      DR      InterPro; IPR003961; FN_III.
CC      DR      InterPro; IPR007110; IG_III.
CC      DR      InterPro; IPR003598; IG_C2.
CC      DR      Pfam; PF00041; Ims; 2.
CC      DR      Pfam; PF00047; Ig; 5.
CC      DR      SMART; SMO0060; FN3; 2.
CC      DR      SMART; SMO0060; IGC2; 5.
CC      DR      PROSITE; PSS0835; IG_LIKE; 5.
CC      DR      Immunoglobulin domain, Cell adhesion; Glycoprotein; Repeat; Signal;
CC      KM      GPI-anchor; Alternative splicing.
CC      FT      SIGNAL 1 19
CC      FT      CHAIN 20 761
CC      FT      DOMAIN 20 111
CC      FT      DOMAIN 116 205
CC      FT      DOMAIN 212 301
CC      FT      DOMAIN 308 403
CC      FT      DOMAIN 406 491
CC      FT      DOMAIN 518 595
CC      FT      DOMAIN 660 727
CC      FT      DISULFID 41 96
CC      FT      DISULFID 139 189
CC      FT      DISULFID 235 287
CC      FT      DISULFID 329 385
CC      FT      DISULFID 426 479
CC      FT      CARBOHYD 222 232
CC      FT      CARBOHYD 315 315
CC      FT      CARBOHYD 347 347
CC      FT      CARBOHYD 423 423
CC      FT      CARBOHYD 449 449
CC      FT      CARBOHYD 478 478
CC      FT      VARSPPLIC 655
CC      FT      NALMK (1n isoform C).
CC      FT      /FTid=VSP_002587.
CC      SQ      SEQUENCE 761 AA; 83770 MM; P0CAD392D7A867E CRC64;
CC      Query Match 6.2%; Score 145; DB 1; Length 761;
CC      Best Local Similarity 20.5%; Pred. No. 0.02;
CC      Matches 100; Conservative 67; Mismatches 211; Indels 110; Gaps 20;
CC      3 RG-VPRHLLLVQLALPAATQGNKRVLGKKGVDELCTTASQKSKSIQFHKNSNQIKI 61
CC      DB 196 RGEINFKDIOIVNVPTTIGARQNIYVNAVNTANIGSGSTVLVCADEGFPEPTMSW----- 247
```

```
QY      62 LGNGSFLTKGPKSLNDRADSRRLMDGNFPLIITKLIKIEDSPYICEVE-----DQKEE 117
DB      248 -----TKDGEIQEGEDDEKTIIFSDSSQLTIKKYKNDDEAETICIAENKAGEQDAT 299
QY      118 VQLLVFG-----LTANSPTHLLOQSLTLTLLESPPSSPSVQCR-----SPRGNIQCG- 166
DB      300 IHLKVFAPKRTIYVENQFAMELEBOVTLTCEASGDPFSITWRTSTRNISSEKTLDDGM 359
QY      167 -----KTLVSQLEIQDSGTWTCTVLQ--NOKYEFKIDIVVLAFOKASIVYKKEG 216
DB      360 VVRSHARVSSLTLLKSITQYTDAGEYICTASNTIGDQSGMYLEVQYAPKLOGPVAVYTWEG 419
QY      217 EGYE-----PSPPLAFYTEKLTGSGELMWOAERASSSSKSWITPDLK-KEVSVKRVTDGP 270
DB      420 NQVNTTCVFVFPFH-TIS-----WFRDGLPSSNYSNITKIYTPASISYEVTIDS 470
QY      271 KLQNGK-----KLPLHLTLPOA-----LPOYAGSGNLTALBAKTKGLHQ 310
DB      471 ENDGNYNCTAVNRHIGESLEFLVQADTPSSPIDQVEPYSSTAQVQFDEPEATG----- 526
QY      311 EVNLVWRATOLQKVLTEWVGPTSPKMLSLKLENK--EAKVSKREKPVVW--LNPBAG 366
DB      527 --GVPILKYKAEWPAVGEVW-----HSKWDYDAKESMEGIVTIVGLKPEPTT 571
QY      367 MMQCLSDSGQVLE-----SNIKVLPWTSTPVHPRASALPAPPGSALPDQGTASALPDP 422
DB      572 YAVRIALANGKGLGELISASAEFTQPVHSP--PPASASSSTPVPLSPDP-----TTWPLP 625
QY      423 PAASALPA 430
DB      626 ALATTEPA 633
```

```
RESULT 19
ID      VCAM1_MOUSE STANDARD; PRT; 739 AA.
AC      P29513;
DT      01-APR-1993 (Rel. 25, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Vascular cell adhesion protein 1 precursor (V-CAM 1).
GN      VCAM1 OR VCAM-1.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORM 1).
RC      STRAIN=FVB; TISSUE=Lung;
RX      MEDLINE=92181437; Pubmed=1371918;
RA      Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wyck M.,
RA      Burkly L., Miyake K., Kincade P., Lobb R.;
RT      "Cloning of murine and rat vascular cell adhesion molecule-1.";
RL      Biochem. Biophys. Res. Commun. 183:163-169(1992).
RN      [2]
RP      SEQUENCE FROM N.A. (ISOFORM 1).
RC      TISSUE=Lymph node;
RX      MEDLINE=93246254; Pubmed=7683304;
RA      Araki M., Araki K., Vassalli P.;
RT      "Cloning and sequencing of mouse VCAM-1 cDNA.";
RL      Gene 126:261-264(1993).
RN      [3]
RP      SEQUENCE FROM N.A. (ISOFORM 1).
RC      STRAIN=129; TISSUE=Embryo;
RX      MEDLINE=94117008; Pubmed=7507076;
RA      Cybulsky M.I., Allan-Motamed M., Collins T.;
RT      "Structure of the murine VCAM1 gene.";
RL      Genomics 18:387-391(1993).
RN      [4]
RP      SEQUENCE OF 1-693 FROM N.A. (ISOFORM 1).
RC      STRAIN=NIH Swiss; and 129/Sv;
RA      Kumar A.G., Dai Y.X., Kozak C.A., Mims M.P., Gotto A.M. Jr.,
```



```

Db 271 SGNATLTLIAMEDESGVYCEGVNLIGRDKAEVLVQEKPIVDISFGQVAAQVDS 330
Oy 139 LTLTESPPGSSPVQOCSPRGNKIQ-----DKTISVSOLELDQSDTWTCTYLQNK 191
Db 331 VVLTCAGCGDSPSPSWRTQDTPINGVNRGAKSTLVSSGFEDESHYCAATCLOK 390
Oy 192 KVEFKIDIVLAFQKASISIVYKKEGQVFEFPLA-----FTVEKL 232
Db 391 TLEKRTQVEVVSF-----PEDPVIMSGPVLHGRPVYVNCVPPVPRDHEIELL 441
Oy 233 TGSGLMWAQERASSKSMI-TFDLKNNEVSVKRVTDQKQ-WGKKL---PLHLTLPO 286
Db 442 KC-----ETTLMKKYFLEMGGIKSLKLETKLETFPTIDTGSILVCLARLHSGME 493
Oy 287 ALPOYAGSGNLTALAEAKTGKLGHOENLVYMRATQLOK---NLTCVWGFPSPLMLSL 342
Db 494 SPPKQROSQVPLVNVAP-----KETTIVWSSPILBEGSPVNLTCSSDGIAPAKILMSR 548
Oy 343 KLENKAKVSKREKPPWVLN---PEAGMWC 370
Db 549 QANNELOPLSENTLTFTWSTRDSDGIYVC 579

RESULT 20
NEOL MOUSE STANDARD; PRT: 1493 AA.
ID NEOL MOUSE
AC P97758;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neogenin precursor.
GN NEOL OR NGN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Brain;
RX MEDLINE=97407661; PubMed=9264410;
RA Keeling S.L., Gad J.M., Cooper H.M.;
RT "Mouse neogenin, a DCC-like molecule, has four splice variants and is
RT expressed widely in the adult mouse and during embryogenesis.";
RL Oncogene 15:691-700(1997).
CC -1- FUNCTION: May be involved as a regulatory protein in the
CC transiation of undifferentiated proliferating cells to their
CC differentiated state. May also function as a cell adhesion
CC molecule in a broad spectrum of embryonic and adult tissues.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=P97798-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P97798-2; Sequence=VSP_002594;
CC Name=3;
CC IsoId=P97798-3; Sequence=VSP_002595;
CC Name=4;
CC IsoId=P97798-4; Sequence=VSP_002596;
CC Note=Expression developmentally regulated;
CC Name=5;
CC Note=Expression developmentally regulated;
CC IsoId=P97798-5; Sequence=VSP_002597;
CC Note=Expression developmentally regulated;
CC -1- TISSUE SPECIFICITY: Widely expressed.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED UBICITOUSLY THROUGHOUT THE MID TO
CC LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION
CC IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE
CC E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF
CC THE MESENCEPHALON AND RHOMENCEPHALON. ISOFORMS 3 AND 4 ARE
CC EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5
CC AND E16.5.

```

```

CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily, DCC family.
CC -1- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 6 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y09535; CAA70727.1; -.
CC HSSP: P02751; 1TPE.
CC MGD: MGI:1097159; Neol.
CC InterPro: IPR008957; FN_III-like.
CC InterPro: IPR003961; FN_III.
CC InterPro: IPR003962; FNIII_subd.
CC InterPro: IPR007110; IG_1like.
CC InterPro: IPR003598; IG_c2.
CC Pfam: PF00041; fn3; 6.
CC Pfam: PF00047; Ig; 4.
CC PRINTS: PR00014; FNTYPEIII.
CC SMART: SM00060; FN3; 6.
CC SMART: SM00408; IGc2; 4.
CC PROSITE: PS50835; IG_LIKE; 4.
CC Cell adhesion, Repeat, Signal, Transmembrane; Immunoglobulin domain;
CC Glycoprotein; Alternative splicing.
CC SIGNAL 1
CC CHAIN 37 1493
CC DOMAIN 37 1136
CC TRANSMEM 1137 1157
CC DOMAIN 1158 1493
CC DOMAIN 63 158
CC DOMAIN 163 249
CC DOMAIN 254 347
CC DOMAIN 352 437
CC DOMAIN 467 564
CC DOMAIN 567 660
CC DOMAIN 661 760
CC DOMAIN 766 860
CC DOMAIN 881 981
CC DOMAIN 982 1083
CC DOMAIN 1149 1153
CC DISULFD 85 140
CC DISULFD 184 232
CC DISULFD 281 331
CC DISULFD 373 421
CC CARBOHYD 84 84
CC CARBOHYD 221 221
CC CARBOHYD 337 337
CC CARBOHYD 501 501
CC CARBOHYD 520 520
CC CARBOHYD 670 670
CC CARBOHYD 746 746
CC CARBOHYD 940 940
CC VARSPLIC 442 461
CC FT VARSPLIC 863 878
CC FT VARSPLIC 1086 1096
CC FT VARSPLIC 1279 1331
CC SO SEQUENCE 1493 AA; 163159 MW; 441DE919D5E17C0B CRC64;
Query Match 6.1%; Score 141.5; DB 1; Length 1493;
Best Local Similarity 23.2%; Pred. No. 0.083;
Matches 117; Conservative 54; Mismatches 188; Indels 145; Gaps 28;
Oy 10 LLLVLTALL--PAA-----TQGNKV-----VLGKKDPTVELT 40
Db 25 LTLTLPLLLLGRPASGAATKSGPRQSGASVRTFTPFVFLVPEVDTLSVRGSSVLIN 84

```


RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O. ;
 RT "Prediction of the coding sequences of unidentified human genes. XIII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro." ;
 RN DNA Res. 6:63-70(1999).
 RN (9)
 RN REVISIONS.
 RC TISSUE=Brain;
 RX MEDLINE=22156633; PubMed=12168954;
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T. ;
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual
 RT curation of 330 KIAA cDNA clones." ;
 RL DNA Res. 9:99-106(2002).
 RN (10)
 RP SEQUENCE OF 5754-6885 FROM N.A.
 RX MEDLINE=21154917; PubMed=11230166;
 RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glaser S.,
 RA Ansgorge W., Boecher M., Bloecher H., Bauersachs S., Blum N.,
 RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
 RA Mewes H.-W., Oltersbinder B., Obermaier B., Tampe J., Heubner D.,
 RA Wandt R., Korn B., Klein M., Poustka A. ;
 RT "Towards a catalog of human genes and proteins: sequencing and
 RT analysis of 500 novel complete protein coding human cDNAs." ;
 RL Genome Res. 11:422-435(2001).
 CC -1- FUNCTION: Involved in the maintenance of nuclear organization and
 CC structural integrity. Probably anchoring protein which tethers the
 CC nucleus to the cytoskeleton. Connects nuclei to the cytoskeleton
 CC by interacting with the nuclear envelope and with F-actin in the
 CC cytoplasm.
 CC -1- SUBUNIT: Interacts with F-actin via its N-terminal domain.
 CC -1- SUBCELLULAR LOCATION: Type IV membrane protein (Potential). The
 CC largest part of the protein is cytoplasmic, while its C-terminal
 CC part is associated with the nuclear envelope, most probably the
 CC outer nuclear membrane. Remains associated with the nuclear
 CC envelope during its breakdown in mitotic cells.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=9;
 CC Name=1;
 CC IsoId=Q8WXH0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8WXH0-2; Sequence=VSP_007164, VSP_007166;
 CC Note=No experimental confirmation available;
 CC Name=3;
 CC IsoId=Q8WXH0-3; Sequence=VSP_007155;
 CC Note=Produced by exon skipping that results in a frameshift. No
 CC experimental confirmation available;
 CC Name=4; Synonyms=Beta;
 CC IsoId=Q8WXH0-4; Sequence=VSP_007156;
 CC Name=5; Synonyms=Alpha;
 CC IsoId=Q8WXH0-5; Sequence=VSP_007157, VSP_007164, VSP_007165;
 CC Name=6;
 CC IsoId=Q8WXH0-6; Sequence=VSP_007158, VSP_007165, VSP_007166;
 CC Note=No experimental confirmation available;
 CC Name=7; Synonyms=Gamma;
 CC IsoId=Q8WXH0-7; Sequence=VSP_007154, VSP_007163;
 CC Name=8;
 CC IsoId=Q8WXH0-8; Sequence=VSP_007161, VSP_007162;
 CC Note=No experimental confirmation available;
 CC Name=9; Synonyms=NUNANCE-N-33;
 CC IsoId=Q8WXH0-9; Sequence=VSP_007159, VSP_007160;
 CC -1- TISSUE SPECIFICITY: Widely expressed, with higher level in kidney,
 CC adult and fetal liver, stomach and placenta. Weakly expressed in
 CC skeletal muscle and brain. Isoform 5 is highly expressed in
 CC pancreas, skeletal muscle and heart.
 CC -1- DOMAIN: The Klarsicht domain mediates the nuclear envelope
 CC targeting.
 CC -1- SIMILARITY: Belongs to the neegrin family.
 CC -1- SIMILARITY: Contains 1 actin-binding domain.
 CC -1- SIMILARITY: Contains 2 calponin-homology (CH) domains.
 CC -1- SIMILARITY: Contains 1 Klarsicht domain.
 CC -1- SIMILARITY: Contains 13 leucine-rich (LRR) repeats.
 CC -1- SIMILARITY: Contains 9 spectrin repeats.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF435010; AALJ3547.1; -
 DR EMBL; AF435011; AALJ3548.1; -
 DR EMBL; AY061757; AALJ3800.1; -
 DR EMBL; AY061758; AALJ3801.1; -
 DR EMBL; AY061759; AALJ3802.1; -
 DR EMBL; AF495911; AAN60443.1; -
 DR EMBL; AL117404; CAB55905.1; -
 DR EMBL; AL162832; -; NOT_ANNOTATED_CDS.
 DR EMBL; ALJ35094; -; NOT_ANNOTATED_CDS.
 CC -----
 Query Match 6.1%; Score 141; DB 1; Length 6885;
 Best Local Similarity 25.6%; Pred. No. 0.65;
 Matches 83; Conservative 59; Mismatches 128; Indels 54; Gaps 15;
 QY 96 IKMKIEDSDTYICEVEDQKEEVLVGLTA--NSDTHLQ--GQSITLTLESPPGSSP 151
 DB :::::--LRACEETKKEIKEVPEETLAQNNLEHATLNBAQNFVLEVSNDVVGSSI 647
 QY 152 SVQCRSPKQKIQGKITSVQLEI-----ODSGTW--TCYVLQNK--VEFKDI 199
 DB :::::--LNRKRWKLVSKQLEMLNPLMKQDQPFDSNGILSKKEATVEFSTDM 704
 QY 200 VVLAFOKASIV-----YKKEGOVEFSPFLAFYVEKLTSGSELWQNERASSKSWITP 254
 DB :::::--LRAKQVNVKQIAALKNLT-----DVSPDLDIRKME--SQKELSYMRAQQLLG 869
 QY 705 SVLEPENTNQIKAGEKHEKENESEFTQGLAKAVKEKIGVEI--WEAE---AKSVLDQ 759
 DB :::::--LKNKEVSVKRV---TDDPKLQMGKLPVHLTPQALPOYAGSGNLTALFAKTKGL 308
 QY 255 D--LKNKEVSVKRV---TDDPKLQMGKLPVHLTPQALPOYAGSGNLTALFAKTKGL 308
 DB DDDVTSMESESKHLIAKSMFDELMARESDMLQMDIQISSQSEFQVHVTGLQAKIGA 819
 QY 309 HOEVNVLVVRATQLOKNTTCVWGPTSPKMLSLKLEKKAQVSKREKPVAVLNPDAQM 368
 DB :::::--LSDSGGVLTLESNIKVL 387
 QY 369 QC-----LSDSGGVLTLESNIKVL 387
 DB 870 QRESPELISKXKHALISNTKSL 893
 DB -----
 RESULT 22
 DCC MOUSE STANDARD; PRT; 1447 AA.
 AC P70211;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Tumor suppressor protein DCC precursor.
 GN DCC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOPFORMS A, B AND C).
 RC STRAIN=BALB/c; TISSUE=Brain;
 RX MEDLINE=96112625; PubMed=8570174;
 RA Cooper H.M., Armes P., Britto J., Gad J., Wilks A.F. ;
 RT "Cloning of the mouse homologue of the deleted in colorectal cancer
 RT gene (mdcc) and its expression in the developing mouse embryo." ;
 RL Oncogene 11:2243-2254(1995).
 RN [2]
 RP REVISIONS.
 RC STRAIN=BALB/c; TISSUE=Brain;
 RA Cooper H.M. ;
 RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.

```

CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=A;
CC IsoId=P70211-1; Sequence=Displayed;
CC Note=Isoform B is produced by alternative initiation at Met-85
CC of isoform A;
CC Name=C;
CC IsoId=P70211-2; Sequence=VSP_002501;
CC Event-Alternative initiation;
CC Comment=2 isoforms, A (shown here) and B, are produced by
CC alternative initiation at Met-1 and Met-85;
CC -1- TISSUE SPECIFICITY: In the embryo, expressed at high levels in the
CC developing brain and neural tube. In adult, highly expressed in
CC brain with very low levels found in testis, heart and thymus.
CC Isoform C is expressed only in the embryo.
CC -1- DEVELOPMENTAL STAGE: Low levels in early gestation. Highest levels
CC expressed during mid gestation. Levels decrease in late gestation
CC and remain at this level in the adult.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. DCC family.
CC -1- SIMILARITY: Contains 6 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X85788; CAA59786.1; -.
DR HSP; P56276; ITLK.
DR MGD; MGI:94869; DCC.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII subd.
DR InterPro; IPR007110; IG_1-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00041; fn3; 6.
DR Pfam; PF00047; Ig; 4.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SMO0060; FN3; 6.
DR SMART; SMO0408; IGc2; 3.
DR PROSITE; PS50835; IG_LIKE; 4.
KM Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KM Repeat; Anti-oncogene; Alternative initiation; Alternative splicing.
FT SIGNAL 1 25
FT CHAIN 26 1447 TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM A.
FT INIT MET 85 85 FOR ISOFORM B.
FT DOMAIN 26 1097 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1098 1122 POTENTIAL.
FT DOMAIN 1123 1447 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 136 135 IG-LIKE C2-TYPE 1.
FT DOMAIN 139 229 IG-LIKE C2-TYPE 2.
FT DOMAIN 234 326 IG-LIKE C2-TYPE 3.
FT DOMAIN 331 416 IG-LIKE C2-TYPE 4.
FT DOMAIN 426 522 FIBRONECTIN TYPE-III 1.
FT DOMAIN 525 618 FIBRONECTIN TYPE-III 2.
FT DOMAIN 619 716 FIBRONECTIN TYPE-III 3.
FT DOMAIN 722 816 FIBRONECTIN TYPE-III 4.
FT DOMAIN 840 940 FIBRONECTIN TYPE-III 5.
FT DOMAIN 941 1042 FIBRONECTIN TYPE-III 6.
FT DISULFID 61 117 BY SIMILARITY.
FT DISULFID 161 212 BY SIMILARITY.
FT DISULFID 261 310 BY SIMILARITY.
FT DISULFID 352 400 BY SIMILARITY.
FT CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 819 838 Missing (in isoform C).
FT SEQUENCE 1447 AA; 158298 MW; 0D1F1097C2D25B9F CRC64;
SQ
Query Match 6.0%; Score 139; DB 1; Length 1447;
Best Local Similarity 20.9%; Pred. No. 0.12;
Matches 76; Conservative 48; Mismatches 131; Indels 108; Gaps 14;
QY 34 GDVLELTCTASQKKSIOFMKNSNQIKILGNQGSFLTKGPSKUND-RADSRRLMDQGNF 92
DB 154 GDTVLTKCEVIGBPMFTIHWQKNOQ-----DINPLPGDSRRVVLPBG-- 195
QY 93 PLIIKLIKIDSDTYICEVD-----QKEVQLL-----VRELTANSDTHLLOQ 137
DB 196 ALQISRLQPDSDSVYCSARNPASIRTGNAEVRITSDPGLHRLQVFLQRPNSVIALEGK 255
QY 138 SLTL---TLESPGGS-----SPSVOCSPRGKNIQGGKTVLSOLELQSGTWCTVLQ 188
DB 256 DAVLECCVSGYPPSPFTWLRGEEVQLRSK-YSLQSGNLSNVTDDSDGTYTCVYTY 314
QY 189 NQKKVEFKIDIVLA---FQKASSIYKKEGQVSPFLAFTVEKLTGSGELMWOAERA 245
DB 315 KNNISASAEITVLPWPMFLNHPNSNLVAYESMDIEF-----ECA 353
QY 246 SSSKSWTTP-LNKKVSVKRVQDPKLGKPLHLTLPOALPOYAGSGNLT----- 299
DB 354 VSGKPVFTVMMKNGDV-----VPSYFQVGGSNRLILGVK 392
QY 300 -----ALEATGKLGHEVNLVWRATQKNTLCEVWGPSPKMLSLKLENKAKV 351
DB 393 SDGFPYOCVAVENAGNAGNSAQLIVPKPAIPSSIL-----PSAPRVPLVSSRVRL 447
QY 352 SKR 354
DB 448 SWR 450
RESULT 23
NCM2_MOUSE STANDARD; PRT; 837 AA.
AC O35136; O35962;
ID 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neutral cell adhesion molecule 2 precursor (N-CAM 2) (RB-8 neural cell)
DE adhesion molecule) (R4B12).
GN NCM2 OR OCAM OR RNCAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC STRAIN=BALB/c; TISSUE=Olfactory neuroepithelium;
RX MEDLINE=97368238; Pubmed=9221781;
RA Yoshitara Y., Kawasaki M., Tamada A., Fujita H., Hayashi H.,
RA Kagamiyama H., Mori K.;
RT "OCAM: A new member of the neural cell adhesion molecule family
RT related to zone-to-zone projection of olfactory and vomeronasal
RT axons.";
RL J. Neurosci. 17:5830-5842(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC STRAIN=C57BL/6J; TISSUE=Olfactory epithelium;
RX MEDLINE=97476194; Pubmed=9334170;
RA Alenius M., Bohm S.;
RT "Identification of a novel neural cell adhesion molecule-related gene
RT with a potential role in selective axonal projection.";
RL J. Biol. Chem. 272:26083-26086(1997).
CC -1- FUNCTION: May play important roles in selective fasciculation and

```

zone-to-zone projection of the primary olfactory axons.

-1- SUBCELLULAR LOCATION: Type I membrane protein (long isoform) and attached to the membrane by a GPI-anchor (short isoform).

-1- ALTERNATIVE PRODUCTS:

Event:Alternative splicing; Named isoforms=2;

 Name:Long;
 IsoId=O55136-1; Sequence=Displayed;

 Name:Short;
 IsoId=O55136-2; Sequence=VSP_002590;

-1- TISSUE SPECIFICITY: Expressed in subsets of both olfactory and vomeronasal neurons in a zone-specific manner.

-1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.

-1- SIMILARITY: Contains 2 fibronectin type III domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sb.ch/announce/or_send_an_email_to_license@isb-sb.ch).

CC EMBL; AF001287; AAB69125.1; -;
CC EMBL; AF001286; AAB69124.1; -;
CC EMBL; AF016619; AAC53375.1; -;
CC MGD; MGI:97282; Ncam2.
CC InterPro; IPR008957; FN_III-like.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR007110; IG_1like.
CC InterPro; IPR003598; IG_C2.
CC Pfam; PF00041; fn3; 2.
CC Pfam; PF00047; Ig; 5.
CC SMART; SMO0060; FN3; 2.
CC SMART; SMO0408; IGC2; 5.
CC PROSITE; PS00835; IG_LIKE; 5.
KW Cell adhesion; Transmembrane;
KW Immunoglobulin domain; Signal; GPI-anchor; Alternative splicing.
KW SIGNAL

FT SIGNAL 1
FT CHAIN 19
FT DOMAIN 20 837
FT TRANSMEM 20 697
FT DOMAIN 719 837
FT DOMAIN 21 108
FT DOMAIN 113 202
FT DOMAIN 208 297
FT DOMAIN 302 396
FT DOMAIN 401 491
FT DOMAIN 482 581
FT DOMAIN 594 678
FT DISULFID 42 93
FT DISULFID 136 186
FT DISULFID 232 281
FT DISULFID 322 380
FT DISULFID 422 475
FT CARBOHYD 177 177
FT CARBOHYD 219 219
FT CARBOHYD 309 309
FT CARBOHYD 406 406
FT CARBOHYD 419 419
FT CARBOHYD 445 445
FT CARBOHYD 474 474
FT CARBOHYD 562 562
FT CARBOHYD 694 837
FT VARSPLIC

SEQUENCE 837 AA; 93203 MW; 70473B053A2D65A5 CRC64;

Query March 5 9%; Score 138; DB 1; Length 837;
Beat local Similarity 19.2%; Pred. No. 0.066;
Matches 93; Conservative 74; Mismatches 190; Indels 128; Gaps 22

QY	3	IG-VEPRHLHLVQLALPRAATQSKNVLGKGDVELLTCTASQKKSIGPHMNSQIYI	61
Db	193	RGEIDFRIIIVVNPVPAIMPMQKSFMAVAEGEBEETLCTKASGSPDPTISPRNG--KL	250
QY	62	LGNGSFLTKGSPSKINDRADSRRLMDQGNFPLIIKNLKIEDSDTYICEV-----EDQKE	116
Db	251	IENKRYILKG-----SNTELYRNIIINKOGGSYVCATNAGEDQK	293
QY	117	E-VQLLVFGLTANSDTLLQGS-----LTLLESPPGSSPSVQC-----	155
Db	294	AFLAQVAV-----PHILQLKNETTSENGHVLVCEAGEBPVEITWRAIDGVMFSEG	346
QY	156	-RSPRG-----KNIOGKTLVSQQLDLOSGTWTYQLQ-----NOKKVERKIDLVLAFOK	206
Db	347	DKSPDGRILEVKGQHRSSLDHIDVKLSBGRYDCENASRIGHQGRMHLDIETAP-KFVS	405
QY	207	ASSIYYKEGEGEOVEFSFPLAFVTEKLTGSGELMWAERASSKSWITFDLKNKEVSVKRV	266
Db	406	NOTWYMSWEGNPNINISCDVT-----ANPPASLHMWREKLL-----LPAKNT	446
QY	267	TODPFLQWKKLPLHLTLPLQALPOYAGSGNLTLLAEKGTGLHOEVNLVWMPATOLQKL	326
Db	447	TLKHTSGYGRKWLIEIA-PTSDNDF--GRVNCIATNIRIGTFQOEYILIELADV-----	495
QY	327	TCEWVGSPSPKMLSLKLENKAKYSKKEP-----VWVLNDEAGWMOCLSD	374
Db	496	-----PSSPHGVCKITIELSQTAKTIS-ENKPSHSGGVPIHHYQVDVKEVASETWKIVRSH	548
QY	375	SGQVU-----LESNIKVLPTWSTPPHPRASALPAPPTG--SALPDPQTASAL-PDEPAAS	426
Db	549	GVQTMVNVLSLEPN-----TTEIRPAAVNGKGQGGQYSKIEIRQTLPVNEPSPSIIH	600
QY	427	ALPAA 431	
Db	601	GQPS 605	

RESULT 24

CEAS_HUMAN	STANDARD;	PRT;	702 AA.
AC	P06731;		
DT	01-JAN-1988 (Rel. 06, Created)		
DT	01-DEC-1992 (Rel. 24, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Carcinoembryonic antigen-related cell adhesion molecule 5 precursor		
DE	(Carcinoembryonic antigen) (CEA) (Meconium antigen 100) (CD66e		
DE	antigen) .		
GN	CEACAM5 OR CEA.		
OS	Homo sapiens (Human) .		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Buteria; Primates; Carnivora; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=90258861; PubMed=2342461;		
RA	Schrewe H., Thompson J., Bona M., Hefta L.J.F., Manuya A.,		
RA	Haseauer M., Shively J.E., von Kliest S., Zimmermann W.,		
RT	"Cloning of the complete gene for carcinoembryonic antigen: analysis		
RT	of its promoter indicates a region conveying cell type-specific		
RT	expression.";		
RL	Mol. Cell. Biol. 10:2738-2748(1990) .		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=86038676; PubMed=3670312;		
RA	Beauchemin N., Benchimol S., Courroyer D., Fuks A., Stannera C.P.,		
RT	"Isolation and characterization of full-length functional cDNA clones		
RT	for human carcinoembryonic antigen.";		
RL	Mol. Cell. Biol. 7:3221-3230(1987) .		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=89122014; PubMed=3220478;		
RA	Barnett T., Goebel S.J., Nothdurft W.A., Eling J.J.;		

RT	"Carcinoembryonic antigen family characterization of cDNAs coding for NCA and CEA and suggestion of nonrandom sequence variation in their conserved loop-domains"; Genomics 3:59-66(1988).
RN	[4]
RP	SEQUENCE OF 5-702 FROM N.A.
RX	MEDLINE=87128144; PubMed=3814146;
RA	Oikawa S., Nakazato H., Kosaki G.;
RT	"Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequence."/
RL	Biochem. Biophys. Res. Commun. 142:511-518(1987).
RM	[5]
RP	SEQUENCE OF 331-702 FROM N.A.
RX	MEDLINE=87204247; PubMed=3036371;
RA	Zimmermann W., Ottlieb B., Friedrich R., von Kleist S.;
RT	"Isolation and characterization of cDNA clones encoding the human carcinoembryonic antigen reveal a highly conserved repeating structure.";
RL	Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC	-1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC	-1- TISSUE SPECIFICITY: Found in adenocarcinomas of endoderally derived digestive system epithelium and fetal colon.
CC	-1- PTM: COMPLEX IMMUNOREACTIVE GLYCOPROTEIN WITH A MW OF 180 kDa COMPRISING 6x CARBOHYDRATE.
CC	-1- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC	-1- SIMILARITY: Contains 7 immunoglobulin-like domains.
CC	-1- DATABASE: NAME=PROW; NOTE=CD guide CD66 entry;
CC	MIM="http://www.ncbi.nlm.nih.gov/prow/cd/cd66e.htm".
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/ or send an email to license@isb.slb.ch).
CC	-----
DR	EMBL, M17303; AA859513.1; .
DR	EMBL, M59262; AAA62835.1; ALT SEQ.
DR	EMBL, M59255; AAA62835.1; JOINED.
DR	EMBL, M59256; AAA62835.1; JOINED.
DR	EMBL, M59257; AAA62835.1; JOINED.
DR	EMBL, M59258; AAA62835.1; JOINED.
DR	EMBL, M59259; AAA62835.1; JOINED.
DR	EMBL, M59260; AAA62835.1; JOINED.
DR	EMBL, M59261; AAA62835.1; JOINED.
DR	EMBL, M59709; -. NOT ANNOTATED CBS.
DR	EMBL, M59710; -. NOT ANNOTATED CBS.
DR	EMBL, M29540; AA51967.1; .
DR	EMBL, X16455; CA93474.1; .
DR	EMBL, M15042; AA51963.1; .
DR	EMBL, M16234; AA51972.1; .
DR	PIR, A36319; A36319.
DR	PDB, 1E07; 04-JUL-00.
DR	GeneID: HGNC:1817; CEACAM5.
DR	MIM, 114890; .
DR	GO; GO:0005887; C: integral to plasma membrane, TAS.
DR	InterPro; IPR007110; IG-like.
DR	Pfam; PF00047; Ig; 6.
DR	PROSITE; PS00835; IG_LIKE; 6.
KW	Immunoglobulin domain; Glycoprotein; Lipoprotein; GPI-anchor; Membrane; Signal; Repeat; 3D-structure.
FT	SIGNAL
FT	1 34
FT	CHAIN 35 685
FT	-----
FT	PROPEP 686 702
FT	DOMAIN 35 144
FT	DOMAIN 146 237
FT	DOMAIN 238 322
FT	DOMAIN 324 415
FT	DOMAIN 416 498
FT	DOMAIN 502 593
FT	DOMAIN 594 677
FT	-----
FT	CARCINOEMBRYONIC ANTIGEN-RELATED CELL ADHESION MOLECULE 5.
FT	REMOVED IN MATURE FORM (POTENTIAL). -----

Query Match	Best Local Similarity	5.9% Score 137.5; DB 1; Length 702;	Matches	96; Conservative	18.8%; Pred. No. 0.059;	76; Mismatches	215; Indels	123; Gaps	20
FT LIPID	685	685							
FT CARBOHYD	104	104							
FT CARBOHYD	115	115							
FT CARBOHYD	152	152							
FT CARBOHYD	182	182							
FT CARBOHYD	197	197							
FT CARBOHYD	204	204							
FT CARBOHYD	208	208							
FT CARBOHYD	246	246							
FT CARBOHYD	256	256							
FT CARBOHYD	274	274							
FT CARBOHYD	288	288							
FT CARBOHYD	292	292							
FT CARBOHYD	309	309							
FT CARBOHYD	330	330							
FT CARBOHYD	351	351							
FT CARBOHYD	360	360							
FT CARBOHYD	375	375							
FT CARBOHYD	432	432							
FT CARBOHYD	466	466							
FT CARBOHYD	480	480							
FT CARBOHYD	508	508							
FT CARBOHYD	529	529							
FT CARBOHYD	553	553							
FT CARBOHYD	560	560							
FT CARBOHYD	580	580							
FT CARBOHYD	612	612							
FT CARBOHYD	650	650							
FT CARBOHYD	665	665							
FT CONFLICT	320	320							
SEQUENCE	702 AA;	76795 MM;	62999AE26CDDDBB5C CRC64;						
Query Match	5.9%;	Score 137.5; DB 1; Length 702;							
Best Local Similarity	18.8%;	Pred. No. 0.059;							
Matches	96; Conservative	76; Mismatches	215; Indels	123; Gaps	20				
20	PAATQGNKRVIVGKKGDVVELCTGASQKKSIOFHWNKSNQIKILGNQGSFLLTKGPKSLNDR	79							
147	PSISSNNKPEVDK-DAVAFTCEBETODATYALMVNNOSLPV-----	187							
80	ADRSRLMDQGNFPLIKLKLIEDSDTYICEVED-----QKEBYOL-LVFGLTANS----	129							
188	--SPPLQSLNSNRRTLLTFNVRNDPTASTKYCCETQNVASRRSDSVTLNVLVYGPDAFTISPL	245							
130	DTHLLQGSLLTLTLTPSPGSSPSVQCR--SPRGKNIQGKTLVSLSOLELQSDGTTWTCVLQ	188							
246	NTSVRSRSGNMLLSCHA--ASNPPAQYSFMVNGTPOQSQELFIPITVNNSSGYTQCAHN	303							
189	NQKKV-EPKIDIVLAFQKASSIYYKKEGEQYERSFPLATVE-KLTGSGELMWAQERAS	246							
304	SDTGINRTVTITVYAEPPKPFITSNNSNEVEDEDAVALTCPEPIQNTVYLLMVNNOSL	363							
247	SSKSWITFDLKNKEKSVKRVQDP-----KLQWGGKLLPLH-----TLPOA	287							
364	PVSPRLQLSNNRRTLLTILSTRNDVGPKEGCIQNELSDHSDPVLNVLVYGPDDPTISPS	423							
288	LPOYAGSNTLLALEAKT-----GKLQOEY-NLVVWRATQLOKNT-TCEWGGPT	334							
424	YTYRPGVNLISLCSAASNPPAQYSWMLDGINIQHTQELFTISNTEKNSGLYTCOANNSA	483							
335	SPKMLSLKLENKRAKV-----SKREPV-----WVLPDEAGMQLCLSDSGOVLLE	381							
484	SGHSRTVYKTTVSSELKPSPSISSNSKSPVEDKDAVAFTCEPEAONTTYLLMVNNQO----	539							
382	SNIKVLPTWSTVPHPRASALPAPPT-----GSALPEPOTASAL	419							
540	-----SLPVEPRQLQSLNGNRTLLTFNTRNDARAVYCGIQNSVGSANRSDPVTLDVL	590							
420	--PDPPAASALPALAVISFLLGLGLVAC	447							
591	YGPDPITISPPDS-----SYLSGNLNLNSC	615							

```

RESULT 25
CAML_RAT
ID CAML_RAT STANDARD; PRT; 1259 AA.
AC 005695;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neural cell adhesion molecule L1 precursor (N-CAM L1).
GN L1CAM OR CAML1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISORFMS 1 AND 2).
RX MEDLINE=91372414; PUBMED=1894011;
RA Miura M., Kobayashi M., Asou H., Uemura K.;
RT "Molecular cloning of cDNA encoding the rat neural cell adhesion
RT molecule L1. Two L1 isoforms in the cytoplasmic region are produced
RT by differential splicing.";
RL FEBS Lett. 289:91-95(1991).
CC -1- FUNCTION: Cell adhesion molecule with an important role in the
CC development of the nervous system. Involved in neuron-neuron
CC adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds
CC to axonin on neurons.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: Named isoforms=2;
CC Name=1;
CC IsoId=Q05695-1; Sequence=Displayed;
CC Name=2; Synonyms=L1cs;
CC IsoId=Q05695-2; Sequence=VSP_002592;
CC -1- TISSUE SPECIFICITY: Isoform 2 is predominantly found in the brain,
CC while isoform 1 is found in the peripheral nervous system.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC L1/neurofascin/NCAM family.
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 5 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, X59149; CAA1860.1; -.
DR PIR, S36126; S36126.
DR HSSP, P20241; 1CPB.
DR InterPro, IPR008957; FN_III-like.
DR InterPro, IPR003961; FN_III.
DR InterPro, IPR007110; IG_1-like.
DR InterPro, IPR003598; IG_C2.
DR Pfam, PF00041; fn3; 4.
DR Pfam, PF00047; Ig_6; 6.
DR SMART, SM00060; FN3; 4.
DR SMART, SM00408; IG2; 5.
DR PROSITE, PS00835; IG_LIKE; 6.
DR Neurogenesis; Cell adhesion; Developmental protein; Glycoprotein;
DR Transmembrane; Repeat; Immunoglobulin domain; Signal;
KW Alternative splicing;
KW SIGNAL
FT CHAIN 1 19 BY SIMILARITY.
FT CHAIN 20 1259 NEURAL CELL ADHESION MOLECULE L1.
FT DOMAIN 1123 1145 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1146 1259 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 135 128 IG-LIKE C2-TYPE 1.
FT DOMAIN 138 128 IG-LIKE C2-TYPE 2.
FT DOMAIN 239 327 IG-LIKE C2-TYPE 3.
FT DOMAIN 332 419 IG-LIKE C2-TYPE 4.
FT DOMAIN 424 506 IG-LIKE C2-TYPE 5.
FT DOMAIN 517 600 IG-LIKE C2-TYPE 6.

```

```

FT DOMAIN 827 896 FIBRONECTIN TYPE-III 1.
FT DOMAIN 932 994 FIBRONECTIN TYPE-III 2.
FT DOMAIN 1032 1093 FIBRONECTIN TYPE-III 3.
FT SITE 553 555 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 562 564 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 57 113 BY SIMILARITY.
FT DISULFID 157 218 BY SIMILARITY.
FT DISULFID 263 311 BY SIMILARITY.
FT DISULFID 353 403 BY SIMILARITY.
FT DISULFID 447 496 BY SIMILARITY.
FT DISULFID 538 590 BY SIMILARITY.
FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 824 824 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 875 875 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 968 968 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1021 1021 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1029 1029 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1072 1072 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1106 1106 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1179 1182 Missing (in isoform 2).
SQ SEQUENCE 1259 AA; 140934 MW; 0F12A7E4415F3C08 CRC64;
Query Match 5.8%; Score 135; DB 1; Length 1259;
Best Local Similarity 20.2%; Pred. No. 0.19;
Matches 66; Conservative 58; Mismatches 114; Indels 88; Gaps 13;
QY 10 LLLVQLALPLPAQGNKRVLGKGDYELTCTISQKKSIO-FHWKNSNOIKILGNQGS 67
DB 507 ILAMIQVEATQIOTGPRSTIEKKGARVTFQASFPDSIQASTRWDRG----- 557
QY 68 FLTGPSKLNDRADSRSLMDQGNPLIKLKIEDSDTYC-----EVEDQKEVQLLVF 123
DB 558 -----DLQGRGSDKFTIEDGQ-LVTKSLDVSQGDVSCVASTLDEVESRAQLLVV 608
QY 124 GLTAN-----SDTHLQGSILTLTLESPGSSPVQCRSP-----RGKNIQGGKTL 169
DB 609 GSPGPVPHLESDHLLKQSGVHLSW-----SPAEDHNSPIETKYDIEFEDKEMAPEKMF 662
QY 170 SVSQLELDQSGTWTCTYLQONQKVEFKIDIVL-----AFQKASSI 210
DB 663 SLGKV-----PQNQSTTLKLSPLYHYFRVAINKYGGEPSPVSEITVPEAPPEKNPV 718
QY 211 VYKKGEGQ-----VERSP-LATFVE-KLTGSGELMMQARA-----SSSK 249
DB 719 DVREGNETNMVITWKPELRMMDMNAPQIQIRVQMRPLKQETWKEQTVSDPFLVSVTS 778
QY 250 SWITFDLKNKEVSRYVTDPKLQMG 275
DB 779 TFVPEYELVQAVNNQGGKPEPQVITG 804

```

```

RESULT 26
SHS1_RAT
ID SHS1_RAT STANDARD; PRT; 509 AA.
AC P97710; 008951; 070426; 09QW15;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein-tyrosine phosphatase non-receptor type substrate 1 precursor
DE (SHP substrate-1) (SHPS-1) (Inhibitory receptor SHPS-1) (Signal-
DE regulatory protein alpha-1) (S1rp-alpha-1) (Brain Ig-like molecule

```


Query Match	Similarity	Score	Length	509
Beat Local	Similarity	22.0%	Pred. No. 062	
Matches	Conservative	75	Mismatches 150	Indels 127
			Gaps	24
23	TOGNKIVLGGKGDVETLTCTTASQKSI----	OFHMKNSNQIKILNQSGSFLTKGSPSKLND	78	
Db	37	TOADSKSVAAAGDSATLNCVTSSLPVGPRIKFKFGQGRSPDIYSFIDGEHPRITNVSD	96	
Qy	79	RADSRSLMDQGNFPLIKLKIKIESDPTICEVEQKEVOLLVGLTANSTHLLQGS	138	
Db	97	-ATKNNM---DFGISISNVPREDAGTYCC-VKROKQIVE-----POTEIKSGGG	141	
Qy	139	LTLTSPSPGSPSVQCSRPAGKNIQGGKTLSSVQLDLDGDTWTCTVLQNKVEPKID	198	
Db	142	TLTYLVLAAP-SSEPVSQSPRSK---SPQGVN-----FTC-----KSYGFS--	178	
Qy	199	IVVLAFOKASSIVYKKEGEQVAFSPFLATVEKLTGSGELMMQABRASSKSMITFDLKN	258	
Db	179	-----PRNITLKMVKDGKELS-----HLEFTTSKSNVSYNI-S	211	
Qy	259	KEYSVKRYATQDPKLOMGKKLPYLHLPALPQYASGNLTALF-AKTGKLMQEVNLYVM	317	
Db	212	STVSVKLSPEEDHSRIVEV-AHVL-EGRP-LNGTANFMSNIIRASPLTIKIQO--PLT	265	
Qy	318	RATQLOKQLTCGEVWGPSTPKMLSLKLEKAKVSPREK-----VVVILN	362	
Db	266	PASQV-NLTQCV-QKTYPKAIQLMWLEN-GNLSRTDREHFTTNRDGTATYVLSLFLVN	320	
Qy	363	PEAG-----MMQSLSDSGQVLLESNIKVLPWTSTVHNPASALPAPRTSALPDDQTSAA	418	
Db	321	SSAHNEDEVVFTQGVHDSQPAITENHTV-----RAFAHSSSGSM-----	360	
Qy	419	LDPDPAASALPAAALVISFLGLGACVCL	449	

ID	CAML_MOUSE	STANDARD;	PRT;	1260 AA.
RESULT 27				
DR	CAML_MOUSE			
AC	P11627.			
DT	01-OCT-1989 (Rel. 12, Created)			
DT	01-OCT-1989 (Rel. 12, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Neural cell adhesion molecule L1 precursor (N-CAM L1).			
GN	L1CAM OR CAML1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	TISSUE=Brain;			
RX	MEDLINE=88315924; PubMed=3412448;			
RA	Moos M., Tacke R., Scherer H., Teplow D., Fruh K., Schachner M.;			
RT	"Neural adhesion molecule L1 as a member of the immunoglobulin			
RT	superfamily with binding domains similar to fibronectin.";			
RT	Nature 334:701-703(1988).			
CC	-1- FUNCTION: Cell adhesion molecule with an important role in the			
CC	development of the nervous system. Involved in neuron-neuron			
CC	adhesion, neurite fasciculation, outgrowth of neurons, etc. Binds			
CC	to axonin on neurons.			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-1- SIMILARITY: Belongs to the immunoglobulin superfamily.			
CC	L1/neurofascin/NG2CAM family.			
CC	-1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.			
CC	-1- SIMILARITY: Contains 5 fibronectin type III domains.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-sib.ch).			
CC	or send an email to license@isb-sib.ch .			
CC	-----			
DR	EMBL; X12875; CA331368.1; -.			
DR	PIR; S05479; S05479.			
DR	HSSP; P20241; 1CFB.			
DR	MGP; MGI:96721; L1cam.			
DR	GO; GO:0007411; P-axon guidance; IMP.			
DR	InterPro; IPR008957; FN_III-like.			
DR	InterPro; IPR003961; FN_III.			
DR	InterPro; IPR007110; IG_1like.			
DR	InterPro; IPR003598; IG_c2.			
DR	Pfam; PF00041; fn3; 4.			
DR	Pfam; PF00041; fn3; 6.			
DR	SMART; SM00060; FN3; 2.			
DR	SMART; SM00408; IGc2; 5.			
DR	PROSITE; PS50835; IG_LIKE; 6.			
KW	Neurogenesis; Cell adhesion; Developmental protein; Glycoprotein;			
KW	Transmembrane; Repeat; Immunoglobulin domain; Signal.			
FT	SIGNAL	1	19	
FT	CHAIN	20	1260	NEURAL CELL ADHESION MOLECULE L1.
FT	DOMAIN	20	1123	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1124	1146	POTENTIAL.
FT	DOMAIN	1147	1260	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	35	130	IG-LIKE C2-TYPE 1.
FT	DOMAIN	138	225	IG-LIKE C2-TYPE 2.
FT	DOMAIN	239	327	IG-LIKE C2-TYPE 3.
FT	DOMAIN	332	419	IG-LIKE C2-TYPE 4.
FT	DOMAIN	424	506	IG-LIKE C2-TYPE 5.
FT	DOMAIN	517	600	IG-LIKE C2-TYPE 6.
FT	DOMAIN	827	896	FIBONECTIN TYPE-III 1.
FT	DOMAIN	932	994	FIBONECTIN TYPE-III 2.
FT	DOMAIN	1032	1094	FIBONECTIN TYPE-III 3.
FT	SITE	553	555	CELL ATTACHMENT SITE (POTENTIAL).

```

FT SITE 562 564 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 57 113 BY SIMILARITY.
FT DISULFID 157 208 BY SIMILARITY.
FT DISULFID 263 311 BY SIMILARITY.
FT DISULFID 353 403 BY SIMILARITY.
FT DISULFID 447 496 BY SIMILARITY.
FT DISULFID 538 590 BY SIMILARITY.
FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 824 824 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 875 875 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 968 968 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 978 978 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1022 1022 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1073 1073 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1107 1107 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1260 AA; 140968 MM; 22B57001CG2A538 CRG64;

```

```

Query Match 5.8%; Score 134.5; DB 1; Length 1260;
Best Local Similarity 20.1%; Pred. No. 0.2;
Matches 73; Conservative 59; Mismatches 140; Indels 91; Gaps 15;

```

```

QY 34 GDTVELTCTAQSQKSIQFHMKNSNIXILNGSGFLTGSPSKNDRAISRSLMDQGNFP 93
DB 346 GETARLDQVQVGRPOPEITWR-----INGSMETVNVKQCKYRI-EQGS-- 387
QY 94 LIINKLKIEDSDTYICEVEDQK-----EEVOLVFGTLFANSDDL-LQGOSILTLTL 143
DB 388 LILSVQCTDTMNVNOCSEARNQHGLLANAYIVVQLPARILTKNQNTMAVEGSAVYLIC 447
QY 144 ESPPSSSVQVQCRSGRKNII-----QGKTLVSQLEIQQDSGTWTCTVLQNKQYEF 195
DB 448 KAFGAPVSVQWLDDEGTTVLQDERFFPYANGTSLIRDLQANDTGRYFCQANQNNVTI 507
QY 196 KIDIVYLAFOKASSI-----VYKKEGROVESFPLAFTVEKTLGSGSLMMQARASSK 249
DB 508 LANLQV---KEATQITQGPRAIETKKGARVTFTCQASFDPSL--QASITWRGDR--- 557
QY 250 SWITFDLKNKEYSVGRVTQDPFLQWCKLPLHLTPQALPOYAGSGNLTLLAEAKTKLH 309
DB 558 -----DLQERGSDSKYFIEDGLVI-----QSL-DYSDQGNVSCVASFELDEVE 600
QY 310 QEVNLVNRK-----TQIQKNTCTGVWGP---TSPKMLSLKLENKEAK 350
DB 601 SRAQLLVVSGPVPVPHLELSDRLHLKQCVHLS---WSPADHNSPIEKYIEDKEMA 657
QY 351 VSK 353
DB 658 PEK 660

```

RESULT 28

```

CAML_HUMAN
ID CAML_HUMAN STANDARD; PRT; 1257 AA.
AC P32004; Q8TA87;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neural cell adhesion molecule L1 precursor (N-CAM L1) (CD117 antigen).
GN L1CAM OR CAML1 OR NC15.
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92031698; PubMed=1932117;
RA Kobayashi M., Mura M., Asou H., Uyemura K.;
RT "Molecular cloning of cell adhesion molecule L1 from human nervous
RT tissue: a comparison of the primary sequences of L1 molecules of
RT different origin.";
RL Biochim. Biophys. Acta 1090:238-240(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92120663; PubMed=1769655;
RA Hlavin M.L., Lemmon V.;
RT "Molecular structure and functional testing of human L1CAM: an
RT interspecies comparison.";
RL Genomics 11:416-423(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92329299; PubMed=1627459;
RA Reid R.A., Hemperly J.J.;
RT "Variants of human L1 cell adhesion molecule arise through alternate
RT splicing of RNA.";
RL J. Mol. Neurosci. 3:127-135(1992).
RN [4]
RP SEQUENCE FROM N.A.
RA Rosenthal A., Coutelle O., Drescher B.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=97432815; PubMed=9286695;
RA Brenner V., Nyakatura G., Rosenthal A., Platzner M.;
RT "Genomic organization of two novel genes on human Xq28: compact head
RT to head arrangement of IDH gamma and TRAP delta is conserved in rat
RT and mouse.";
RL Genomics 44:8-14(1997).
RN [6]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=98147998; PubMed=9479034;
RA Coutelle O., Nyakatura G., Taudien S., Elgar G., Brenner S.,
RA Platzner M., Drescher B., Jonek M., Kenwright S., Rosenthal A.;
RT "The neural cell adhesion molecule L1: genomic organization and
RT differential splicing is conserved between man and the pufferfish
RT Fugu.";
RL Gene 208:7-15(1998).
RN [7]
RP SEQUENCE OF 20-36.
RX MEDLINE=88298876; PubMed=3136168;
RA Wolff J.M., Frank R., Mujoo K., Spiro R.C., Reisfeld R.A.,
RA Rathjen F.G.;
RT "A human brain glycoprotein related to the mouse cell adhesion
RT molecule L1.";
RL J. Biol. Chem. 263:11943-11947(1988).
RN [8]
RP SEQUENCE OF 332-371 FROM N.A.
RX MEDLINE=90353957; PubMed=2387585;
RA Djahali M., Matei M.-G., Nguyen C., Roux D., Demengeot J.,
RA Denizot F., Moos M., Schachner M., Goridis C., Jordan B.R.;
RT "The gene encoding L1, a neural adhesion molecule of the
RT immunoglobulin family, is located on the X chromosome in mouse and
RT man.";
RL Genomics 7:587-593(1990).
RN [9]
RP SEQUENCE OF 353-1176 FROM N.A.
RC TISSUE=fetal brain;
RX MEDLINE=92020233; PubMed=1923824;
RA Rosenthal A., Mackinnon R.N., Jones D.S.C.;
RT "PCR walking from microdissection clone M54 identifies three exons
RT from the human gene for the neural cell adhesion molecule L1
RT (CAM-L1)."
RL Nucleic Acids Res. 19:5395-5401(1991).

```



```

Db 687 FRVTAINKYGPSPVSEVTVPTEPAEAKNPVDVKGEGNET-----TNWVI 733
Oy 317 MRATOLQKNTCEWGP-----TSPKMLSLKLENKEAKVSKKEKPYWVLPNPAQMWQ- 369
Db 734 -----TWKPLRMMDNNAFQVQYRVQ-----WRQGTGRGMOE 765
Oy 370 CLUSDGQVLLSNIKVLPTWSTPVPAPASALPAPPTGSALPDPTASALPDPPASALP 429
Db 766 QIVSD--PFLVVSNTSTP-----VPYEIKVQAVNSQSGK---PEQVTVIGSGEDYPPAIP 816

RESULT 29
KMLS_CHICK STANDARD; PRT; 1906 AA.
AC P11799; P19038;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myosin light chain kinase, smooth muscle and non-muscle isozymes
DE (EC 2.7.1.117) (MCK) [Contains: Telokin].
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM MCK-210).
RX MEDLINE=96033976; PubMed=7589469;
RA Waterston D.M., Collinge M., Lukas T.J., van Eldik L.J., Birukov K.G.,
RA Stepanova O.V., Shitinsky V.P.;
RT "Multiple gene products are produced from a novel protein kinase
RT transactivation region.";
RT FEBS Lett. 373:217-220(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM MCK-108).
RX MEDLINE=90192792; PubMed=2315320;
RA Olson N.J., Pearson R.B., Needleman D.S., Hurwitz M.J., Kemp B.E.,
RA Means A.R.;
RT "Regulatory and structural motifs of chicken gizzard myosin light
RT chain kinase.";
RT Proc. Natl. Acad. Sci. U.S.A. 87:2284-2288(1990).
RN [3]
RP SEQUENCE OF 649-1906 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Fibroblast;
RX MEDLINE=90361738; PubMed=2202734;
RA Shoemaker M.O., Lau W., Shatlock R.L., Kwiatkowski A.P.,
RA Marriestian P.E., Guerra-Santos L., Wilson E., Lukas T.J.,
RA van Eldik L.J., Waterston D.M.;
RT "Use of DNA sequence and mutant analyses and antisense
RT oligodeoxynucleotides to examine the molecular basis of nonmuscle
RT myosin light chain kinase autoinhibition, calmodulin recognition, and
RT activity.";
RT J. Cell Biol. 111:1107-1125(1990).
RN [4]
RP SEQUENCE OF 1259-1906 FROM N.A.
RC TISSUE=Gizzard;
RX MEDLINE=87157587; PubMed=3030394;
RA Guerriero V. Jr., Russo M.A., Olson N.J., Putkey J.A., Means A.R.;
RT "Domain organization of chicken gizzard myosin light chain kinase
RT deduced from a cloned cDNA.";
RT Biochemistry 25:8372-8381(1986).
RN [5]
RP SEQUENCE OF 1750-1906 FROM N.A. (ISOFORM TELOKIN).
RC TISSUE=Gizzard;
RX MEDLINE=93073972; PubMed=1444462;
RA Yoshikai S., Ikebe M.;
RT "Molecular cloning of the chicken gizzard telokin gene and cDNA.";
RT Arch. Biochem. Biophys. 299:242-247(1992).
RN [6]
RP SEQUENCE OF 1750-1906 FROM N.A. (ISOFORM TELOKIN).
RX MEDLINE=92236611; PubMed=1373815;
RA Collinge M., Marriestian P.E., Zimmer W.E., Shattuck R.L., Lukas T.J.,
RA van Eldik L.J., Waterston D.M.;

```

```

RT "Structure and expression of a calcium-binding protein gene contained
RT within a calmodulin-regulated protein kinase gene.";
RL Mol. Cell. Biol. 12:2359-2371(1992).
CC -1- FUNCTION: Phosphorylates a specific serine in the N-terminus of a
CC myosin light chain, which leads to the formation calmodulin/MCK
CC signal transduction complexes which allow selective transduction
CC of calcium signals.
CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
CC light-chain] phosphate.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative Initiation;
CC MCK-108/Smooth-muscle and Telokin, are produced by alternative
CC initiation;
CC -1- TISSUE SPECIFICITY: ISOFORM TELOKIN IS EXPRESSED IN GIZZARD,
CC HEART, LUNG, INTESTINE, AND SKELETAL MUSCLE ALTHOUGH THE LEVELS OF
CC THE EXPRESSION IN THE LATTER WERE MUCH LESS THAN THAT IN THE
CC GIZZARD.
CC -1- DOMAIN: TELOKIN BINDS CALMODULIN.
CC -1- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC -1- SIMILARITY: Contains 9 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DR EMBL; X52876; CA37056.1; -
DR EMBL; X52876; CA37057.1; -
DR EMBL; X52876; CA37058.1; -
DR EMBL; M31048; AAA49089.1; -
DR EMBL; M14953; AAA69964.1; -
DR EMBL; M96655; AAA49083.1; -
DR EMBL; M88283; AAA48647.1; -
DR EMBL; M88284; AAB53768.1; -
DR PIR; S68235; S68235.
DR PDB; 1CDL; 31-ARG-94.
DR PDB; 1VRK; 27-APR-99.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG_1-like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR007119; Prot_kinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; ig; 9.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00408; IGC2; 8.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PSS0835; IG_LIKE; 9.
DR PROSITE; PSS0011; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PSS00108; PROTEIN_KINASE_DOM; 1.
DR Transfaser; Serine/threonine-protein kinase; Calmodulin-binding;
KW ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
KW Alternative initiation; 3D-structure.
FT CHAIN 1 1906
FT CHAIN 935 1906
FT CHAIN 1750 1906
FT CHAIN 1750 1906
FT INT_MET 935 935
FT INT_MET 1750 1750
FT DOMAIN 28 117
FT DOMAIN 156 244

```

```

FT DOMAIN 429 517 IG-LIKE C2-TYPE 3.
FT DOMAIN 521 613 IG-LIKE C2-TYPE 4.
FT DOMAIN 637 725 IG-LIKE C2-TYPE 5.
FT DOMAIN 735 830 IG-LIKE C2-TYPE 6.
FT DOMAIN 1084 1172 IG-LIKE C2-TYPE 7.
FT DOMAIN 1225 1313 IG-LIKE C2-TYPE 8.
FT DOMAIN 1330 1400 FIBRONECTIN TYPE-III.
FT DOMAIN 1453 1708 PROTEIN KINASE.
FT DOMAIN 1794 1885 IG-LIKE C2-TYPE 9.
FT NP BIND 1457 1467 ATP (BY SIMILARITY).
FT BINDING 1462 1462 ATP (BY SIMILARITY).
FT ACT SITE 1574 1574 BY SIMILARITY.
FT DOMAIN 1716 1728 CALMODULIN AUTOINHIBITION (AM13) REGION
FT FT (POTENTIAL).
FT FT CALMODULIN RECOGNITION (RS20) REGION
FT FT (POTENTIAL).
FT DOMAIN 1730 1749 MOTIF IA.
FT FT MOTIF IB.
FT DOMAIN 1317 1364 4 X REPEATS, MOTIF IIA.
FT DOMAIN 1385 1402 4 X REPEATS, MOTIF IIA.
FT DOMAIN 660 676 4 X REPEATS, MOTIF IIA.
FT REPEAT 758 774 IIA-1.
FT REPEAT 1107 1123 IIA-2.
FT REPEAT 1817 1833 IIA-3.
FT DOMAIN 693 1866 IIA-4.
FT REPEAT 708 791 5 X REPEATS, MOTIF IIB.
FT REPEAT 791 807 IIB-1.
FT REPEAT 1140 1156 IIB-2.
FT REPEAT 1281 1287 IIB-3.
FT REPEAT 1851 1866 IIB-4.
FT DOMAIN 970 1226 4 X REPEATS, MOTIF III.
FT REPEAT 987 1016 IIB-1.
FT REPEAT 1061 1078 IIB-2.
FT REPEAT 1209 1226 IIB-3.
FT DOMAIN 1700 1763 IIB-4.
FT DOMAIN 1896 1906 CALMODULIN-BINDING.
FT MOD RES 1748 1748 POLY-GLU.
FT MOD RES 1762 1762 PHOSPHORYLATION.
FT CONFLICT 1439 1439 R -> Q (IN REP. 4)
SQ SEQUENCE 1906 AA; 210445 MW; AD7DBA3869E33363 CRC64;

Query Match 5.6%; Score 130; DB 1; Length 1906;
Beet Local Similarity 20.8%; Pred. No. 0.7;
Matches 82; Conservative 58; Mismatches 142; Indels 112; Gaps 15;

QY 89 OGNFLLITKLNKLEISDYTCVED---QKEEVLLVFGLTNSDTHLQCGSLTLTLE 144
DB 81 RGFSLVLKGVQEGDSGKTYTCEANDGVRQVTVETVEG---NS---LKKYSLPSSAK 133
QY 145 SPPG---SSPSVOCR-----SPRGKNIQGGKT----- 168
DB 134 TPQGLSLVPEVHRHSIMGESPPKATKPNRVNVEGQGRSPSKITGRPOPQVTWTKGD 193
QY 169 -----LSVSOLELODSTGTCTVLQONKKEVEFKIDIV----- 201
DB 194 IHLQGNERNRMFEKGIQYLEIQNVQLADAGIYITCVNNSAKKASVSAELTVQGRDKDT 253
QY 202 -----LA-----FOKASSIVYKKEGEQVESFPPLAFVTEKUTSGEELMW 240
DB 254 HAOPLCMPKPTTLTKALEISNDFXQATSNGIAKELKSTEL-VWETKDRLSAKKETFY 312
QY 241 OABERASSSSKSWITFPLKNKEVSVKRVTDPKLQMKKULPLHLTPQALPOVAGSGLTLA 300
DB 313 TSREAKDGKOGNOENANVPLQDSRGTKGPOVLQKTSSTITLQAVVAQDEPPAEPOTFI 372
QY 301 LEAKTGKLEHVENLVVRATQLOKULTECVWGPTSPKMLSLKLEN-----KEAVYSKRE 355
DB 373 RQAEDEBK--RTVQPLMTTTTQENPSLT-----QGVSPR---SRETENRAGVRKSVKEKRE 423
QY 356 KRVWVLNRPAGMWQCLSDSGO-VLLSENIKLP 388
DB 424 -PLGIIPQFESRPSQSLSESGOEIKFKSVKSGKP 456

```

```

RESULT 30
PGBM_MOUSE
ID PGBM_MOUSE STANDARD; PRT; 3707 AA.
AC 005793;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core
DE protein precursor (HSPG) (perlecan) (PLC).
GN HSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=92078153; Pubmed=1744087;
RA Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M.,
RA Yamada Y., Haasell J.R.,
RT "The complete sequence of perlecan, a basement membrane heparan
RT sulfate proteoglycan, reveals extensive similarity with laminin A
RT chain, low density lipoprotein-receptor, and the neural cell adhesion
RT molecule."
RL J. Biol. Chem. 266:22939-22947(1991).
RN [2]
RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89034110; Pubmed=2972708;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,
RA Yamada Y., Haasell J.R.,
RT "Identification of cDNA clones encoding different domains of the
RT basement membrane heparan sulfate proteoglycan."
RL J. Biol. Chem. 263:16379-16387(1988).
CC -1- FUNCTION: This protein is an integral component of basement
CC membranes. It is responsible for the fixed negative electrostatic
CC charge and is involved in the charge-selective ultrafiltration
CC properties. It serves as an attachment substrate for cells.
CC -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC dimers or stellate structures. It interacts with other basement
CC membrane components such as laminin, prolargin and collagen type
CC IV.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: Found in the basement membranes.
CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: Contains 4 LDP-receptor class A domains.
CC -1- SIMILARITY: Contains 11 laminin EGF-like domains.
CC -1- SIMILARITY: Contains 3 laminin IV domains.
CC -1- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 3 laminin G-like domains.
CC -1- SIMILARITY: Contains 1 EGF-like domain.
CC -1- SIMILARITY: Contains 1 SEA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M77174; AA39911.1; -
DR EMBL; J04054; AA39899.1; -
DR EMBL; J04055; AA39912.1; -
DR PIR; S18252; S18252.
DR PDB; 1GL4; 28-NOV-01.
DR MGD; MGI:96257; Hspg2.
DR GO; GO:0005604; C:basement membrane; IDA.
DR GO; GO:0008104; F:protein localization; IMP.
DR InterPro; IPR008985; Cona_1like_1ec_31.
DR InterPro; IPR00742; EGF_2.
DR InterPro; IPR006209; EGF_1like.

```

```

DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF000008; EGF_4.
DR Pfam; PF000047; Ig_15.
DR Pfam; PF00052; Laminin_B_3.
DR Pfam; PF00053; Laminin_EGF_7.
DR Pfam; PF00054; Laminin_G_3.
DR Pfam; PF00057; LDL_recept_a; 4.
DR Pfam; PF01390; SEA; 1.
DR PRINTS; PR00261; LDLRECEPTOR.
DR ProDom; PD001031; Laminin_B_3.
DR SMART; SM00180; EGF_Lam_7.
DR SMART; SM00408; IGC2_14.
DR SMART; SM00281; LamB_3.
DR SMART; SM00282; LamG_3.
DR SMART; SM00192; LDLa_4.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; 8.
DR PROSITE; PS01186; EGF_2; 5.
DR PROSITE; PS50026; EGF_3; 4.
DR PROSITE; PS50835; IG_LIKE; 15.
DR PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR PROSITE; PS01209; LDRA_1; 4.
DR PROSITE; PS50068; LDRA_2; 4.
DR PROSITE; PS50024; SEA; 1.
DR SIGNAL; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW Extracellular matrix; EGF-like domain; 3D-structure.
FT SIGNAL 1 21
FT CHAIN 22 3707
FT DOMAIN 80 194
FT DOMAIN 195 234
FT DOMAIN 281 319
FT DOMAIN 320 359
FT DOMAIN 360 403
FT DOMAIN 404 504
FT DOMAIN 521 530
FT DOMAIN 531 730
FT DOMAIN 731 763
FT DOMAIN 764 813
FT DOMAIN 814 871
FT DOMAIN 879 923
FT DOMAIN 924 933
FT DOMAIN 934 1125
FT DOMAIN 1126 1158
FT DOMAIN 1159 1208
FT DOMAIN 1209 1265
FT DOMAIN 1275 1324
FT DOMAIN 1325 1334
FT DOMAIN 1335 1529
FT DOMAIN 1530 1562
FT DOMAIN 1563 1612
FT DOMAIN 1613 1670
FT DOMAIN 1671 1771
FT DOMAIN 1772 1865
FT DOMAIN 1866 1954
FT DOMAIN 1955 2049
FT DOMAIN 2050 2148
FT DOMAIN 2149 2244
FT DOMAIN 2245 2343
FT DOMAIN 2344 2436
FT DOMAIN 2437 2532
FT DOMAIN 2533 2619
FT DOMAIN 2620 2720
FT DOMAIN 2721 2809
FT DOMAIN 2810 2895

```

```

FT DOMAIN 2896 2980
FT DOMAIN 2984 3162
FT DOMAIN 3163 3241
FT DOMAIN 3245 3425
FT DOMAIN 3518 3705
FT SITE 65 67
FT SITE 71 73
FT SITE 76 78
FT SITE 3615 3617
FT DISULFID 199 212
FT DISULFID 206 225
FT DISULFID 219 234
FT DISULFID 285 297
FT DISULFID 292 310
FT DISULFID 304 319
FT DISULFID 325 337
FT DISULFID 332 350
FT DISULFID 344 359
FT DISULFID 368 381
FT DISULFID 375 394
FT DISULFID 388 403
FT DISULFID 428 479
FT DISULFID 764 773
FT DISULFID 766 780
FT DISULFID 783 792
FT DISULFID 795 811
FT DISULFID 814 829
FT DISULFID 816 839
FT DISULFID 842 851
FT DISULFID 854 869
FT DISULFID 1159 1168
FT DISULFID 1161 1175
FT DISULFID 1178 1187
FT DISULFID 1190 1206
FT DISULFID 1209 1224
FT DISULFID 1211 1234
FT DISULFID 1237 1246
FT DISULFID 1249 1263
FT DISULFID 1275 1287
FT DISULFID 1277 1293
FT DISULFID 1295 1304
FT DISULFID 1307 1322
FT DISULFID 1563 1572
FT DISULFID 1565 1579
FT DISULFID 1582 1591
FT DISULFID 1594 1610
FT DISULFID 1613 1628
FT DISULFID 1615 1638
FT DISULFID 1641 1650
FT DISULFID 1653 1668
FT DISULFID 1792 1839
FT DISULFID 1886 1932
FT DISULFID 1976 2021
FT DISULFID 2073 2118
FT DISULFID 2170 2215
FT DISULFID 2268 2313

```

Query Match 5.5%; Score 127; DB 1; Length 3707;

Best Local Similarity 19.3%; Pred. No. 2; Indels 168; Gaps 22;

Matches 100; Conservative 66; Mismatches 185;

```

OY 32 KKGDTVELTCTASOKSIQFMKNSQDIKILNGSGSLTGPSGLANDRAISRSLMDQGN 91
DB 2447 KMGQDITLBECLSSGSEPRSSPRWTR-----LGI-----PVTLERPMGLAM-----SH 2488
OY 92 PPLIKNKLIKEDSDTYICEVED---OKEEYVLLVFLGLTANSDDLHLDQGSLLVTLTSSPP 147
DB 2489 AMLTIASVKRSDACTYVCAQNALGTAKQKQVELVDGTGVAFGTPVQVVESELTLEA-- 2546
OY 148 GSSPSVQCRS-----PRGKNIOGKTLVSQLELDQSDSGTWCTCVLQNO 190
DB 2547 GHTATLHCATGNGPPTIHWSKLAPLFWOHRIE-GNTLVYIPVAQDSDSGYICNATNSA 2605

```



```

Query Match      5.4%; Score 125.5; DB 1; Length 725;
Best Local Similarity 20.9%; Pred. No. 0.41;
Matches 95; Conservative 64; Mismatches 187; Indels 109; Gaps 21;

QY 3 RG-VEFHLLVLQALLPATQGNKVLGKKGDVETLTASQKSIQFHW-KNSNOIK 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 196 RGEINFKDIQIVVNPPTVQARQSI VNA TANI GGVTLVCDADGPEPTMWTGDGEPI 255
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 ILNGGSLTGTGSPSLNDRAD--SRRLMDGNGFLLIKNLKIEDSDYICEVE----DQ 114
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 256 -----NEEDERSRVSVDSE--VTIRVDKDEAEYVCIENKAGEQ 297
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 115 KEVQLLVFG-----LTASDTHLLQGSLTLTLESPPSSPVOCR-----SPRGNIQ 164
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 298 DASILKFKFAPKRTIYENQGTAMELEGVTLTCEASGDPISITWRTSTNISEEDLD 357
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 165 GG-----KTLVSQLELQDSGTWCTVYLQNKQKVEFKIDIVLAFQKASSI---- 210
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 358 GHMVVSHARVSSLTKSIQYRDAGEVWCTASNTIGDSQSID--LEFQYAPKLQFVA 414
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 211 VYKKEGGEVE-----FSFPLAFTVVKLTGSGELMMQARASSSKSWITFDLN-KEYSVK 264
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 415 VYTWEGNQNVTCEVFAYPSA-TIS-----WFRDQLLPSSVSNIKIYNTPPSASYL 465
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 265 RVTQDPKLTQMGK-----KLPLHLTLPOA-----LPQYAGSGLTLALEAK 304
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 466 EVTPDSEMDPFGNVTAVNRIGQSLERILVQADTPSSPIDRVEPYSTNQVQDEPEA 525
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 305 TGKHLQEVNLVVMRATQLOKNTLTCVWGPSPKMLSLKENKAKSKREKPVWVLNPE 364
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 526 TG-----GVPIILKTKAKEMKSLGESWHT-----WYDAKEANMEGIIVT----IMGLKPE 570
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 365 AGMWCLSDSGOVV-----LESNIKVLPTWSTP 393
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 571 TTYSDRLLALNGKGLGEMQPSSEKTPQVPELSAP 605
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 32
DCC_HUMAN STANDARD; PRT; 1447 AA.
ID_P43146;
AC 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tumor suppressor protein DCC precursor (Colorectal cancer suppressor).
CN DCC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=95011532; PubMed=7926722;
RA Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
RA Vogelstein B.;
RA "The DCC gene product in cellular differentiation and colorectal
RT tumorogenesis";
RL Genes Dev. 8:1174-1183(1994).
RN [2]
RN SEQUENCE OF 1-750 FROM N.A.
RX MEDLINE=90100559; PubMed=2294591;
RA Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA Vogelstein B.;
RA "Identification of a chromosome 18q gene that is altered in
RT colorectal cancers";
RL Science 247:49-56(1990).
RN [3]
RN SEQUENCE OF 107-472 FROM N.A. (SCRAMBLED EXONS).
RX MEDLINE=91121517; PubMed=1991322;
RA Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M.,
RA Oliner J.D., Kinzler K.W., Vogelstein B.;

```

```

RT "Scrambled exons";
RL Cell 64:607-613(1991).
RN [4]
RX GENE STRUCTURE, AND VARIANT CARCINOMA HIS-1375.
RX MEDLINE=94245241; PubMed=8188295;
RA Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,
RA Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;
RT "The DCC gene: structural analysis and mutations in colorectal
RT carcinomas";
RL Genomics 19:525-531(1994).
RN [5]
RN VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
RX MEDLINE=94243823; PubMed=8187090;
RA Miyake S., Nagai K., Yoshino K., Oco M., Endo M., Yuasa Y.;
RT point mutations and allelic deletion of tumor suppressor gene DCC in
RT human esophageal squamous cell carcinomas and their relation to
RT metastasis";
RL Cancer Res. 54:3007-3010(1994).
CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
CC -1- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL
CC NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.
CC -1- DISEASE: Colorectal tumors that lost their capacity to
CC differentiate into mucus producing cells uniformly lack DCC
CC expression. Inactivation of DCC due to allelic deletion and/or
CC point mutations may cause both lymphatic and hematogenous
CC metastasis of esophageal squamous cell carcinomas.
CC -1- SIMILARITY: Belongs to the immunoglobulin-like C2-type family.
CC -1- SIMILARITY: Contains 4 immunoglobulin-like C2-superfamily domains.
CC -1- SIMILARITY: Contains 6 fibronectin type III domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76132; CAAS3735.1; -
DR EMBL; M32292; AAA53751.1; -
DR EMBL; M32286; AAA52174.1; -
DR EMBL; M32288; AAA52175.1; ALT_SEQ.
DR EMBL; M32290; AAA52176.1; -
DR EMBL; M63696; AAA52177.1; -
DR EMBL; M63700; AAA52178.1; -
DR EMBL; M63702; AAA52179.1; -
DR EMBL; M63718; AAA52180.1; -
DR EMBL; M63698; AAA52181.1; -
DR PIR; A54100; A54100.
DR HSSP; P56276; IRLK.
DR Genew; HGNC:2701; DCC.
DR MTM; 120470; -
DR GO; GO:0004888; F:transmembrane receptor activity; TAS.
DR GO; GO:0007409; P:axonogenesis; TAS.
DR GO; GO:0006917; P:induction of apoptosis; TAS.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_subd.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00041; fn3; 6.
DR Pfam; PF00047; Ig; 4.
DR PRINTS; PR00014; FNTYPEIIT.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00408; IGC2; 3.
DR PROSITE; PS50835; IG_LIKE; 4.
KW Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW Repeat; Anti-oncogene; Disease mutation; Polymorphism.
FT SIGNAL 1 25
FT CHAIN 26 1447 TUMOR SUPPRESSOR PROTEIN DCC.
FT DOMAIN 26 1097 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1098 1122 POTENTIAL.

```



```

FT DOMAIN 1123 1447 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 135 IG-LIKE C2-TYPE 1.
FT DOMAIN 139 229 IG-LIKE C2-TYPE 2.
FT DOMAIN 234 326 IG-LIKE C2-TYPE 3.
FT DOMAIN 331 416 IG-LIKE C2-TYPE 4.
FT DOMAIN 426 522 FIBRONECTIN TYPE-III 1.
FT DOMAIN 525 618 FIBRONECTIN TYPE-III 2.
FT DOMAIN 619 716 FIBRONECTIN TYPE-III 3.
FT DOMAIN 722 816 FIBRONECTIN TYPE-III 4.
FT DOMAIN 840 940 FIBRONECTIN TYPE-III 5.
FT DOMAIN 1042 1042 FIBRONECTIN TYPE-III 6.
FT DISULFID 61 117 BY SIMILARITY.
FT DISULFID 161 212 BY SIMILARITY.
FT DISULFID 261 310 BY SIMILARITY.
FT DISULFID 352 400 BY SIMILARITY.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 168 168 M -> T (in oesophageal carcinoma).
FT VARIANT 201 201 /FTid=VAR_003909.
FT VARIANT 201 201 R -> G.
FT VARIANT 1375 1375 /FTid=VAR_003910.
FT VARIANT 1375 1375 P -> H (in a colorectal carcinoma).
FT CONFLICT 138 138 /FTid=VAR_003911.
FT CONFLICT 233 329 MISSING (IN REF. 3).
FT CONFLICT 421 421 MISSING (IN REF. 3).
SQ SEQUENCE 1447 AA; 158456 MW; 4A861276ED0471F CRC64;

```

Query Match Beel Local Similarity 5.4%; Score 125.5; DB 1; Length 1447; Matches 77; Conservative 46; Mismatches 129; Indels 147; Gaps 15;

```

QY 34 GDTVELTCASQKSIQPHWKNNOIKILGNGSFETKPSKLTNDRASRLMDQGNP 93
DB 154 GDTVLKEVEIGEPEPTIHWKNOQ-----DITP-----GDSRVVVLPSG-A 196
QY 94 LIKNLIKIEDSDTYICEVED-----QKEVQL-----VEGLTANSDTHLQSQS 138
DB 197 LQISRLQPDIGIYCSANPASRTGNAEVRIISDPQLHRQLVFLQRPVVVAIEGMD 256
QY 139 LTL---TLSPPGS-----SPSVQCRSPRKNIQGGKLTLSQLELQDSGTWTCTVION 189
DB 257 AVLKCCVSGPPSPFTWLGRBEVQLRSKK-YSLGGSNLTLISNVTDDSGMYTCVVTYK 315
QY 190 OKKVEFKIDIVLA---FOKASSIYKKKEGEVFEFPLAFVEXLTGSGELMWQAEAS 246
DB 316 NENISASAEITVLPWPFLNPSNITYAESMDIER-----ECTVSG-----KPV 359
QY 247 SSKSWITPDKNKEVSVKAVTODPKLQWKKPLHLTLFQALPOVAGSGNLTALAEAKTG 306
DB 360 PTVNM-----MKQGDV-----VIPSDFQIVGSGNLAIR----- 387
QY 307 KLHQEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVVNLNPEAG 366
DB 388 -----IGVKSDEG 396
QY 367 MWQCLSDSGOVLESINIKVLTPTWSPVPRASALPAPP 405
DB 397 FYQCVANENAGNAGNOTSQILIVP---KPAIPSSSVLPSPAP 432

```

```

RESULT 33
PGM HUMAN
ID P98160; Q16287; Q9H3V5; STANDARD; PRT; 4391 AA.
AC P98160; Q16287; Q9H3V5;
DT 01-OCT-1996 (Rel. 34, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Basement membrane-specific heparan sulfate proteoglycan core

```

```

DE protein precursor (HSPG) (Perlecan) (PLC).
GN HSPG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92112994; PubMed=1730768;
RA Kallunki P., Tryggvason K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor.";
RL J. Cell Biol. 116:559-571(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA TISSUE=Colon, and Skin;
RC MEDLINE=92235084; PubMed=1569102;
RA Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor.";
RL J. Biol. Chem. 267:8544-8557(1992).
RN [3]
RP SEQUENCE OF 22-4391 FROM N.A., AND VARIANT SJ81 TYR-1532.
RA MEDLINE=20553141; PubMed=11101850;
RA Nicole S., Davoine C.-S., Topaloglu H., Cattolico L., Barra D.,
RA Beignon P., Ben-Hamida C., Hammouda H., Cruaud C., White P.S.,
RA Samson D., Urtizberea J.A., Lehmann-Horn F., Weissenbach J.,
RA Hentati F., Fontaine B.;
RT "Perlecan, the major proteoglycan of basement membranes, is altered in
RT patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia).";
RL Nat. Genet. 26:480-483(2000).
RN [4]
RP SEQUENCE OF 1016-1470 FROM N.A.
RA TISSUE=Colon;
RC MEDLINE=91365376; PubMed=1679749;
RA Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA Yi H.F., Iozzo R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1.";
RL Genomics 10:673-680(1991).
RN [5]
RP SEQUENCE OF 890-1396 FROM N.A.
RA TISSUE=Fibroblastoma;
RC MEDLINE=92120660; PubMed=1685141;
RA Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA Tryggvason K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to 1p36.1-3p35 and identification of
RT a BamHI restriction fragment length polymorphism.";
RL Genomics 11:389-396(1991).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RA MEDLINE=94052171; PubMed=8234307;
RA Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN [7]
RP CARBOHYDRATE-LINKAGE SITE ASN-2121.
RA MEDLINE=22660472; PubMed=12754519;
RA Zhang H., Li X.-Y., Martin D.B., Aebbersold R.;
RT "Identification and quantification of N-linked glycoproteins using
RT hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL Nat. Biotechnol. 21:660-666(2003).
RN [8]
RP FUNCTION: This protein is an integral component of basement
RN membranes. It is responsible for the fixed negative electrostatic
RN charge and is involved in the charge-selective ultrafiltration
RN properties. It serves as an attachment substrate for cells.

```

-1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in dimers or stellate structures. It interacts with other basement membrane components such as laminin, prolargin and collagen type IV.

-1- SUBCELLULAR LOCATION: Extracellular.

-1- TISSUE SPECIFICITY: Found in the basement membranes.

-1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED AND O-LINKED OLIGOSACCHARIDES.

-1- DISEASE: Defects in HSPG2 are the cause of Schwartz-Jampel syndrome (SJ51) [MIM:255800]; a rare autosomal recessive disorder characterized by permanent myotonia (prolonged failure of muscle relaxation) and skeletal dysplasia, resulting in reduced stature, kyphoscoliosis, bowing of the diaphyses and irregular epiphyses.

-1- SIMILARITY: Contains 4 LDL-receptor class A domains.

-1- SIMILARITY: Contains 11 laminin EGF-like domains.

-1- SIMILARITY: Contains 22 immunoglobulin-like C2-type domains.

-1- SIMILARITY: Contains 3 laminin G-like domains.

-1- SIMILARITY: Contains 1 SEA domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL, X62515, CAA44373.1, -.

EMBL, M85289, AA52700.1, -.

EMBL, AL445795, CAC18534.1, -.

EMBL, M64283, AA52659.1, -.

EMBL, S76436, AAB21121.2, -.

EMBL, L22078, -; NOT_ANNOTATED_CDS.

PIR, A38096, A38096.

HSSP, P00740, IEDM.

Gene, 2DPAGE, P98160, -.

Gene, HGNC:5273, HSPG2.

MIM, 142461, -.

MIM, 255800, -.

InterPro, IPR008985, ConA_like_jec_gl.

InterPro, IPR000742, EGF_2.

InterPro, IPR006209, EGF_like.

InterPro, IPR006210, IEGF.

InterPro, IPR007110, IG-like.

InterPro, IPR003599, IG_c2.

InterPro, IPR003598, IG_c2.

InterPro, IPR003596, IG_v.

InterPro, IPR000034, Laminin_B.

InterPro, IPR002049, Laminin_EGF.

InterPro, IPR001791, Laminin_G.

InterPro, IPR002172, LDL_receptor_A.

InterPro, IPR000082, SEA_domain.

Pfam, PF00008, EGF, 4.

Pfam, PF00047, IG, 22.

Pfam, PF00052, laminin_B, 3.

Pfam, PF00053, laminin_EGF, 7.

Pfam, PF00054, laminin_G, 3.

Pfam, PF00057, LDL_recept, 4.

Pfam, PF01390, SEA, 1.

PRINTS, PR00261, LDLRECEPTOR.

PRODom, PD003031, Laminin_B, 3.

SMART, SM00181, EGF, 15.

SMART, SM00180, EGF_Lam, 12.

SMART, SM00409, IG, 22.

SMART, SM00408, IGc2, 21.

SMART, SM00406, IGV, 7.

SMART, SM00281, LamB, 3.

SMART, SM00282, LamG, 3.

SMART, SM00192, LDLa, 4.

SMART, SM00200, SEA, 1.

PROSITE, PS00022, EGF_1, 9.

DR PROSITE, PS01186, EGF_2, 6.

DR PROSITE, PS50026, EGF_3, 4.

DR PROSITE, PS50835, IG_Like, 22.

DR PROSITE, PS50025, LAM_G DOMAIN, 3.

DR PROSITE, PS01248, LAMININ_TYPE_EGF, 11.

DR PROSITE, PS01209, LDLRA_1, 4.

DR PROSITE, PS50024, SEA, 1.

KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein; Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain; Extracellular matrix; EGF-like domain; Disease mutation.

KW CHAIN 1

FT 22 4391

FT 80 194

FT 198 235

FT 284 320

FT 324 360

FT 367 404

FT 405 504

FT 521 530

FT 531 730

FT 731 763

FT 764 813

FT 814 871

FT 879 923

FT 924 933

FT 934 1125

FT 1126 1208

FT 1159 1258

FT 1209 1265

FT 1275 1324

FT 1325 1334

FT 1335 1529

FT 1530 1562

FT 1563 1612

FT 1613 1670

FT 1677 1771

FT 1772 1865

FT 1866 1955

FT 1956 2051

FT 2052 2151

FT 2152 2244

FT 2245 2340

FT 2341 2436

FT 2437 2533

FT 2534 2629

FT 2630 2726

Query Match 5.4%; Score 125.5; DB 1; Length 4391;

Best Local Similarity 19.9%; Pred. No. 4.2;

Matches 104; Conservative 69; Mismatches 176; Indels 173; Gaps 24;

32 KKGPVLELTCTASQKSIQFMKNSNOIKLNGSEFLTGPCKLNDRADRSRLMDQGN 91

3125 KVGKAVTLKVCVSAAGEPRSSARWTR-----ISSTPAKLEQRT--YGLMD-SH 3167

92 FPLIKKLIKTEIDGTVCYCEVED----QKEVOLLVFEGLTNSDTHLLOGSLTTLTSSPP 147

3168 AVLDISSAKPSDAGTVCLQANMLGTRQKOVEYIVDTGAMAPGAPVOQAEBAEILTYEA-- 3225

148 GSSPSVQC-----RSP-RGNINOGKTLVSOLELSDSGTWTCTVLQNOK 191

3226 GHTATLRCSATGSPARTIHKSKLRPLPMQHRLLEGDTLLIPRVAQDSGGYLICNATSPAG 3285

192 KVEFKDIVLAFQKASIT--YKKEGEQVEFSF-----PLAFVYEK-----TG 234

3286 HAERTIILHVESPPYATTVPEHASVQAGETVQLCLAHGTPPLTFQWSRYGSSLPGRATA 3345

235 SGELMWAERASSKSM-----ITPDLKNKEVSVKRTQPKQMGKKLPLHLTLPLQALP 289

3346 RNEL-LHFERPADBSGRYRCRVTKNKVGSAAEPAQLLVQGP-----PQSLP 3390

```

OY 290 QY---AGSG---NLTLAEAKTKLHQLVNLVVM---RATOL----- 322
DB 3391 ATSIAGSTPTVQVTPQLETKISIGASVEHCAVPSRQGLWFKEGGLPRGHVQDGV 3450
OY 323 -----QKMLTCEVWGSP-----TSPKMLSLKLENKAKV----- 351
DB 3451 LRIONLDSCQGTGYICQAHGPMWKAQASALVIALPVLINIRTSVQVNVGHAVEPEC 3510
OY 352 ----SKREKPVV-----LNP-----EAGMWQCLISDS-----GQVLL 380
DB 3511 IALGDPKPVQWSKVGGHLRPGIIVOSGGVRIAHVELADAGQYRCLTATMAAGTQSHVLL 3570
OY 381 ESNIKVLPMTWSTPVHPRASGLPAPPTGSALPPDGTASALPDP 422
DB 3571 --LVQALPQISMPEVVRV-----PAGSAVFPCLASGTPP 3604

RESULT 34
LACH SCHAM STANDARD; PRT; 349 AA.
AC 026474;
DB 15-JUL-1998 (Rel. 36, Created)
DB 15-JUL-1998 (Rel. 36, Last sequence update)
DB 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lachesin precursor.
GN LAC.
OS Schistocerca americana (American grasshopper).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridoidea;
OC Acrididae; Cyrtacanthacridinae; Schistocerca.
OX NCBI_TaxID=7009;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE=94038693; PubMed=8223276;
RA Karistrom R.O., Wilder L.P., Bastiani M.J.;
RT "Lachesin: an immunoglobulin superfamily protein whose expression
RT correlates with neurogenesis in grasshopper embryos.";
RL Development 118:509-522(1993).
CC -1- FUNCTION: MAY PLAY A ROLE IN EARLY NEURONAL DIFFERENTIATION AND
CC AXON OUTGROWTH.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC ONLY THOSE CELLS THAT BECOME NEUROBLASTS CONTINUE TO EXPRESS IT.
CC EXPRESSED BY NEUROBLASTS, GANGLION MOTHER CELLS AND NEURONS EARLY
CC IN THEIR LIVES, BUT EXPRESSION BECOMES RESTRICTED TO A SUBSET OF
CC NEURONS AS DEVELOPMENT PROGRESSES. EXPRESSED BY SENSORY NEURONS AS
CC THEY DELAMINATE FROM THE BODY WALL ECTODERM. IT IS ALSO PRESENT ON
CC GROWING AXONS OF THE CNS AND PNS AND BECOMES RESTRICTED TO A
CC SUBSET OF AXONS LATER IN DEVELOPMENT.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ON DIFFERENTIATING NEURONAL CELLS
CC FROM THE ONSET OF NEUROGENESIS IN BOTH THE CENTRAL AND PERIPHERAL
CC NERVOUS SYSTEMS.
CC -1- PTM: The N-terminus is blocked.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; LJ3256; AAC37185.1; -.
DB HSSP; P80362; 1MTL.
DB InterPro; IPR007110; Ig-like.
DB InterPro; IPR003598; Ig_C2.
DB Pfam; PF00047; Ig_3.
DB SMART; SM00408; IGC2; 2.
DB PROSITE; PS50835; IG_LIKE; 3.
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW Repeat; Signal; Lipoprotein.

```

```

FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 332 LACHESIN.
FT PROPEP 333 349 REMOVED IN MATURE FORM (POTENTIAL).
FT DOMAIN 22 127 IG-LIKE V-TYPE.
FT DOMAIN 132 218 IG-LIKE C2-TYPE 1.
FT DOMAIN 222 315 IG-LIKE C2-TYPE 2.
FT DISULFID 43 110 POTENTIAL.
FT DISULFID 154 201 POTENTIAL.
FT DISULFID 244 299 POTENTIAL.
FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 332 332 GPI-anchor amidated glycine (Potential).
SQ SEQUENCE 349 AA; 38974 MM; 5F139A48F849689 CRC64;

Query Match 5.4%; Score 125; DB 1; Length 349;
Beet Local Similarity 23.3%; Pred. No. 0.17; Indels 28; Gaps 7;
Matches 62; Conservative 39; Mismatches 137;

OY 34 GDTVELCTTASQKSIQPHW----KNSNQIKILNQGSGFLTKGSPKLNDRASRRSLWDQ 89
DB 36 GGTVELECSVOYADYPTLMKMKVDRNROYDPLPISTGSLIIRSRFALRYDTASS----- 91
OY 90 GNFPILINLKIEDSDTYICEV----EDQKEVQLVFG--LTASDTHLLOGQSITL 141
DB 92 -TYLQIKDIOETDAGFYQCQVILGNKKITAEVLDQVRPPIVSDNSTRLVWSEGOAV 150
OY 142 TLESPPGS--SPSVQGRS-----PRGNIGGKTLTSQSLDSDSGTCTCTVLQNOKKV 193
DB 151 RLECYAGVPPRPVRSWRENNAILPTGGSTIRGVNLKISRGKDRGYTCVAENGVGK 210
OY 194 EFKIDIVILAROKKSIYVKKEGOVERSPFLAFTVEKLTGSGELMWAERASSKSWIT 253
DB 211 ARRNIAVEVERPPVITVPRRPLGQALQYDMDLECHVEVPPPATLWKDEIVLSNQHYS 270
OY 254 --FDLKNKEVSVKRVTDPLQMGK 276
DB 271 ISHFATDEFTDTRVITIERQYQK 296

RESULT 35
DSCA HUMAN STANDARD; PRT; 2012 AA.
AC 060469; 060468;
DB 16-OCT-2001 (Rel. 40, Created)
DB 15-OCT-2001 (Rel. 40, Last sequence update)
DB 10-OCT-2003 (Rel. 42, Last annotation update)
DE Down syndrome cell adhesion molecule precursor (CHD2).
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RA TISSUE=Brain;
RA MEDLINE=98087574; PubMed=9426258;
RA Yamakawa K., Huot Y.-K., Haendelt M.A., Hubert R., Chen X.-N.,
RA Lyons G.E., Korenberg J.R.;
RT "DSCAM: a novel member of the immunoglobulin superfamily maps in a
RT Down syndrome region and is involved in the development of the
RT nervous system.";
RL Hum. Mol. Genet. 7:227-237(1998).
RN [2]
RP SEQUENCE FROM N.A., AND FUNCTION.
RA MEDLINE=20384934; PubMed=10925149;
RA Agatwala K.L., Nakamura S., Tautzumi Y., Yamakawa K.;
RT "Down syndrome cell adhesion molecule DSCAM mediates homophilic
RT intercellular adhesion.";
RL Brain Res. Mol. Brain Res. 79:118-126(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA MEDLINE=20289799; PubMed=10830953;
RA Hatori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Tocioki Y., Choi D.-K., Groner Y.,

```

RA Soeda E., Ohki M., Takaki T., Sakaki Y., Taudien S., Blehschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Paterson D.,
RA Reichwald K., Rump A., Schillnabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Miyayama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordiek G., Hornischer K., Brund P.,
RA Schärfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramer J., Beck A., Klages S., Hennig S., Resselmann L., Dsgand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehmach H., Reimartz K., Yaspo M.-L.,
RT "The DNA sequence of human chromosome 21."
RL Nature 405:311-319 (2000).
CC -I- FUNCTION: CELL ADHESION MOLECULE THAT CAN MEDIATE CATION-
CC INDEPENDENT HOMOPHILIC BINDING ACTIVITY. COULD BE INVOLVED IN
CC NERVOUS SYSTEM DEVELOPMENT.
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE). THE
CC SHORT ISOFORM MAY BE SECRETED.
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long; Synonyms=CHD2-52;
CC IsoId=O60469-1; Sequence=52;Displayed:
CC Name=Short; Synonyms=CHD2-42;
CC IsoId=O60469-2; Sequence=VSP_002502, VSP_002503;
CC -I- TISSUE SPECIFICITY: Primarily expressed in brain.
CC -I- SIMILARITY: Contains 6 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF023440; AAC17967.1; -
DR EMBL; AF023440; AAC17966.1; -
DR EMBL; AF217525; AAF27525.1; -
DR EMBL; AL163283; CAB90464.1; -
DR EMBL; AL163282; CAB90436.1; -
DR EMBL; AL163281; CAB90444.1; -
DR Genew: HGNC:3039; DSCAM.
DR MIM: 602523;
DR GO: GO:0005887; C: integral to plasma membrane; TAS.
DR GO: GO:0005624; C: membrane fraction; TAS.
DR GO: GO:0007155; P: cell adhesion; TAS.
DR GO: GO:0007399; P: neurogenesis; TAS.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003598; Ig_C2.
DR Pfam: PF000471; fn3; 6.
DR Pfam: PF00047; Ig; 9.
DR SMART: SMO0060; FN3; 6.
DR SMART: SMO0408; IGC2; 7.
DR PROSITE: PS50835; IG-Like; 9.
DR Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
DR Transmembrane; Alternative splicing.
KT SIGNAL 1
FT CHAIN 17
FT DOMAIN 18 2012 DOWN SYNDROME CELL ADHESION MOLECULE.
FT TRANSMEM 18 1593 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1596 1616 POTENTIAL.
FT DOMAIN 1617 2012 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 39 129 IG-LIKE C2-TYPE 1.
FT DOMAIN 125 216 IG-LIKE C2-TYPE 2.
FT DOMAIN 225 305 IG-LIKE C2-TYPE 3.
FT DOMAIN 313 401 IG-LIKE C2-TYPE 4.
FT DOMAIN 407 500 IG-LIKE C2-TYPE 5.
FT DOMAIN 504 592 IG-LIKE C2-TYPE 6.
FT DOMAIN 596 685 IG-LIKE C2-TYPE 7.
FT DOMAIN 690 783 IG-LIKE C2-TYPE 8.
FT DOMAIN 787 883 IG-LIKE C2-TYPE 9.
FT DOMAIN 885 972 FIBRONECTIN TYPE-III 1.

FT	DOMAIN	984	1076	FIBONECTIN TYPE-III 2.
FT	DOMAIN	1088	1177	FIBONECTIN TYPE-III 3.
FT	DOMAIN	1189	1273	FIBONECTIN TYPE-III 4.
FT	DOMAIN	1285	1377	IG-LIKE C2-TYPE 10.
FT	DOMAIN	1380	1463	FIBONECTIN TYPE-III 5.
FT	DOMAIN	1477	1562	FIBONECTIN TYPE-III 6.
FT	DISULFID	46	102	BY SIMILARITY.
FT	DISULFID	145	197	BY SIMILARITY.
FT	DISULFID	246	293	BY SIMILARITY.
FT	DISULFID	335	385	BY SIMILARITY.
FT	DISULFID	428	484	BY SIMILARITY.
FT	DISULFID	525	575	BY SIMILARITY.
FT	DISULFID	617	669	BY SIMILARITY.
FT	DISULFID	711	766	BY SIMILARITY.
FT	DISULFID	809	865	BY SIMILARITY.
FT	DISULFID	1307	1359	BY SIMILARITY.
FT	CARBOHYD	28	78	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	78	78	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	470	470	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	487	487	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	512	512	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	556	556	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	658	658	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	666	666	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	710	710	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	748	748	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	795	795	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	924	924	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1142	1142	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1160	1160	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1250	1250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1271	1271	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1341	1341	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1488	1488	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPPLIC	1562	1571	NFATLNDYGS -> KEAARCKEFS (in isoform Short).
FT	VARSPPLIC	1572	2012	/FTId=VSP 002502.
FT	VARSPPLIC	1572	2012	Missing (in isoform Short).
FT	CONFLICT	1893	2012	/FTId=VSP 002503.
FT	CONFLICT	1893	2012	HRPDLHLPLPRLEADSLFNLNGRPGTSGRDLISGQACLEPQK
FT	CONFLICT	1893	2012	SRTLRLPLPLPRLEADSLFNLNGRPGTSGRDLISGQACLEPQK
FT	CONFLICT	1893	2012	EGAEIGLQAKHMSQSGLSDSLDSRGLKGNPNVANKSYTLV ->
FT	CONFLICT	1893	2012	IGQVTSYICLHLLTEWTF (IN REF. 1).
SEQ	SEQUENCE	2012 AA; 222259 MW; 0E33CFB781A08334 CRC64;		
Query Match	Best Local Similarity	5.4%;	Score 125; DB 1; Length 2012;	
Matches	82; Conservative	18.6%;	Pred. No. 1.7; Mismatches 144; Indels 160; Gaps 17	
Db	20	PATQGNVNVVGKKGDVTELCTASQKSIQFHW-KNSNOIKILDNQSGFLTRKSPKLAND	78	
Db	504	PASIRPMKNITAIAGRDYTHICRVAGIYYYSIKYKNSNLPFRNIRQVAFENNGILKLSD	563	
Db	79	RADSRRSIMDGNFPLLIKNIUKIEDSDTYICEVEDQKEEVOVLVGLTANSOTHLLOQS	138	
Db	564	VO-----KEVDEGEYTCNVLIWQPO-----LTSQS	588	
Db	139	LTLTLESP-----GSSSVQCRBPRG-----KNI	163	
Db	589	VHVTIVKVPPIQPEEPFRFSIGQRAVEIPCVVNSGDLPTITTWQKDRPIPSGLGVTTINDI	648	
Db	164	QGGKTLVSQLELDQSGTWTCTVLQNKQKVEFKIDIVILA-----FOKASS	209	
Db	649	DTSSLRKSNLSLWNGNYTICIANNEAAVLEHQSLIYRVPPKFPVQPRPDQGIYKAVI	708	
Db	210	IVYKKEQVEFSSPLAFTEKLT-GSGELMWQERASSSKSWITFDLKNKEVSKV-----	264	
Db	709	LNCSAEG-----YVPVTIWKVFSKAGAVQFO-PIALNGRIQV---LSNGILLIKHIVE	758	
Db	265	-----RYTQDRLQMGKGLPLHLTLPQALPVQVAGSNLTLLAEAKTGKGLHQEVNVLV	316	
Db	759	EDSGYGLCKVANDVGADVSKSMYTLVKIPAMITSP---NTTLAQOQ-----	803	

```

OY 317 MRATOLQKNTCEVWGPTSPKMLSLKLENKEAKVSKREKPY- - - - -VLNPEAGMW 368
DB 804 -----KXEMSTARG-----EKPIYWEKEDRIINPEMARY 835
OY 369 OCLLSDSGQVLLSNIKVLPT 389
DB 836 LVSTKEVGEVVI-STLQILPT 855

RESULT 36
NCAL_MOUSE STANDARD; PRT; 1115 AA.
ID NCAL_MOUSE
AC P13595; Q61949;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neutral cell adhesion molecule 1, 180 kDa isoform precursor (N-CAM 180)
DE (NCAM-180).
GN NCAM1 OR NCAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OC NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM N-CAM 180).
RC STRAIN=C57BL/6;
RX MEDLINE=87246524; Pubmed=3595653;
RA Barthele D., Santoni M.J., Wille W., Ruppert C., Chaix J.C.,
RA Hirsch M.R., Fontecilla-Camps J.C., Goridis C.;
RT "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for
RT a Mr 79,000 polypeptide without a membrane-spanning region.";
RL EMBL J. 6:907-914(1987).
RN [2]
RP SEQUENCE OF 529-1115 FROM N.A. (ISOFORM N-CAM 140).
RC STRAIN=C57BL/6;
RX MEDLINE=88067687; Pubmed=3684567;
RA Santoni M.-J., Barthele D., Barbae J.A., Hirsch M.-R., Steinmetz M.,
RA Goridis C., Wille W.;
RT "Analysis of cDNA clones that code for the transmembrane forms of the
RT mouse neural cell adhesion molecule (NCAM) and are generated by
RT alternative RNA splicing.";
RL Nucleic Acids Res. 15:8621-8641(1987).
RN [3]
RP SEQUENCE OF 642-1115 FROM N.A. (ISOFORM N-CAM 180).
RX MEDLINE=88283628; Pubmed=3396534;
RA Barbae J.A., Chaix J.C., Steinmetz M., Goridis C.;
RT "Differential splicing and alternative polyadenylation generates
RT distinct NCAM transcripts and proteins in the mouse.";
RL EMBL J. 7:625-632(1988).
RN [4]
RP SEQUENCE OF 804-1081 FROM N.A. (ISOFORM N-CAM 180).
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=88247737; Pubmed=2454455;
RA Barthele D., Vopper G., Wille W.;
RT "NCAM-180, the large isoform of the neural cell adhesion molecule of
RT the mouse, is encoded by an alternatively spliced transcript.";
RL Nucleic Acids Res. 16:4217-4225(1988).
RN [5]
RP SEQUENCE OF 702-1115 FROM N.A. (ISOFORM N-CAM 140).
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=89251563; Pubmed=2721486;
RA Santoni M.J., Barthele D., Vopper G., Boned A., Goridis C., Wille M.;
RT "Differential exon usage involving an unusual splicing mechanism
RT generates at least eight types of NCAM cDNA in mouse brain.";
RL EMBL J. 8:385-392(1989).
RN [6]
RP SEQUENCE OF 20-36.
RX MEDLINE=86140120; Pubmed=3512556;
RA Rougon G., Marshak D.R.;
RT "Structural and immunological characterization of the amino-terminal
RT domain of mammalian neural cell adhesion molecules.";
RL J. Biol. Chem. 261:3396-3401(1986).
-!- FUNCTION: This protein is a cell adhesion molecule involved in

```

```

CC neuron-neuron adhesion, neurite fasciculation, outgrowth of
CC neurites, etc.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=N-CAM 180;
CC IsoId=P13595-1; Sequence=Displayed;
CC Name=N-CAM 140;
CC IsoId=P13595-2; Sequence=VSP_002588;
CC Name=N-CAM 120;
CC IsoId=P13594-1; Sequence=External;
CC -!- SIMILARITY: Contains 2 fibronectin-like C2-type domains.
CC -!- SIMILARITY: Contains 2 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X07200; CA30177.1; -.
CC EMBL; Y00051; -. NOT ANNOTATED_CDS.
CC EMBL; X06328; CA29641.1; -.
CC EMBL; X07195; CA30173.1; -.
CC EMBL; X07244; CA30230.1; -.
CC EMBL; X15051; CA33150.1; -.
CC EMBL; X15052; CA33151.1; -.
CC PIR; A29673; IJMSNL.
CC MGI; MGI:97281; Ncam1.
CC InterPro; IPR008957; FN_III-like.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR007110; IG_1like.
CC InterPro; IPR003598; IG_c2.
CC Pfam; PF00041; fn3; 2.
CC Pfam; PF00047; ig; 5.
CC SMART; SM00408; IGc2; 2.
CC SMART; SM00408; IGc2; 5.
CC PROSITE; PS50835; IG_1like; 5.
CC Cell adhesion; Glycoprotein; Transmembrane; Repeat;
CC Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding.
CC SIGNAL 1 19
CC CHAIN 20 1115
CC NEURAL CELL ADHESION MOLECULE 1, 180 kDa
CC ISOFORM.
CC EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC IG-LIKE C2-TYPE 1.
CC IG-LIKE C2-TYPE 2.
CC IG-LIKE C2-TYPE 3.
CC IG-LIKE C2-TYPE 4.
CC IG-LIKE C2-TYPE 5.
CC FIBRONECTIN TYPE-III 1.
CC FIBRONECTIN TYPE-III 2.
CC FIBRONECTIN TYPE-III 2.
CC HEPARIN-BINDING (POTENTIAL).
CC HEPARIN-BINDING (POTENTIAL).
CC PROBABLE.
CC PROBABLE.
CC PROBABLE.
CC PROBABLE.
CC PROBABLE.
CC PROBABLE.
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC Missing (in isoform N-CAM 140).
CC /Ftd=VSP_002588.
CC Query Match 1115 AA; 119351 MW; 2C93DCD474CFCAF CRC64;
CC 5.3%; Score 124; DB 1; Length 1115;

```

	Best Local Similarity	21.9%	Pred.	No. 0.9:	Mismatches	68;	Gaps	15;
OY	Matches	68;	Conservative	47;	Mismatches	128;	Indels	68;
Dd	3	RG-VEPRHLLVLVQLALLPPAATQGNKVVLGKKGDVTVELTCTASQQKSIFQHWM-KNSNOIK	60					
Oy	: :	:	:	:	:	:	:	:
Dd	GGEINFKDIQIVANVPPTVOAQRSITVNATIANIGSVTLVCDDADGPPEPMTSWTKXGBPIE	255						
Oy	61 ILNGSGSLFTKGPSKLNDRAD--SRRLMDQNPFLLIKNLKEIDSDTYICVE----	DQ 114						
Dd	256 -----NEEDBRSSRSSVSDSSE-VTRINDKDDEAEVCIENKAEGD 297							
Oy	115 KEEOVLVFEG-----LTANSDFTHLLOGSSTLTLESPPGSPSPVQCRR-----SFRGNIIQ	164						
Dd	258 DASIHLYKFAPAKKITVVENQTMELEBOYTTLCEASCDPIPETWRTSTRNISSEODLD	357						
Oy	165 GG-----KTLSVSOLETDGSTGTWTCTYLONOKKYEFKIDIVALFQRASSI-----	210						
Dd	358 GHMVVRSHARVASLTIKSIQYDADECYMCTASNTIGQSOSID--LEFOYPKLQGPA	414						
Oy	211 VYKKEGEQE-----FSFPPLATVEKLGGSGELMMQAERASSSKSWTFDLKN-KEYVK	264						
Dd	415 YTTWGNGNVNIITECFVAFPSA-TIS-----WFRDQLLPSSNYSNIKIYNTPSASYL	465						
Oy	265 RVTODPKLOWG 275							
Dd	466 EYTPSENDFFG 476							
RESULT	37							
YTFN_HAEIN	ID_YTFN_HAEIN STANDARD; PRT; 1298 AA.							
AC	057523;							
DT	01-NOV-1997 (Rel. 35, Created)							
DT	01-NOV-1997 (Rel. 35, Last sequence update)							
DT	28-FEB-2003 (Rel. 41, Last annotation update)							
DE	Protein HI0696.							
CN	HI0696							
OC	Haemophilus influenzae.							
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;							
OX	NCB1_TaxId=727;							
R	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=Rd / KM20 / ATCC 51907;							
RX	MEDIIN=95350630; PubMed=7542800;							
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,							
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,							
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., GCoyne J.D.,							
RA	Scott J.D., Shirley R., Liu L.T.I., Glodek A., Kelley J.M.,							
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,							
RA	Usterback T.R., Hanna M.C., Nguyen D.T., Saudke D.M., Brandon R.C.,							
RA	Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,							
RA	Gneshm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,							
RA	Venter J.C.;							
RT	"Whole-genome random sequencing and assembly of Haemophilus influenzae Rd."							
RL	Science 269:496-512(1995).							
RP	(2)							
RP	IDENTIFICATION BY MASS SPECTROMTRY.							
RX	MEDIATE=20137488; PubMed=10675023.							
RA	Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wißf B.,							
RA	Gray C., Fountoulakis M.;							
RT	"Two-dimensional map of the proteome of Haemophilus influenzae.";							
RL	Electrophoresis 21:411-429(2000).							
CC	-! SIMILARITY: STRONG, TO E.COLI YTFN.							
CC	-----							
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -							
CC	between the Swiss Institute of Bioinformatics and the EMBL outpostation -							
CC	the European Bioinformatics institute. There are no restrictions on its use by non-profit institutions as long as their content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce/ or send an email to licenses@isb.sdb.ch).							

CC	EMBL, U3752; AAC2356.1; -.	Score 124; DB 1; Length 1298;
DR	PIR; A64157; A64157.	Best Local Similarity 22.3%; Pred. No. 1.1;
DR	TIGR; H10696; -.	Matches 109; Conservative 58; Mismatches 178; Indels 144; Gaps 22;
DR	InterPro; IPR007452; DUF490.	
DR	Pfam; PF04357; DUF490; 1.	
DR	Complete proteome.	
DR	SEQUENCE 1298 AA; 141222 MW; 2D905DBE6D9329E9 CRC64;	
Query Match	5.3%; Score 124; DB 1; Length 1298;	
Best Local Similarity	22.3%; Pred. No. 1.1;	
Matches 109; Conservative 58; Mismatches 178; Indels 144; Gaps 22;		
QY	7 FRHLIVQLALLPAATGKNVYGGKGDVYELTCTASQKKSIOFHMKNSQIKILNQG	66
DB	815 YRFEKLYIPKLTLMADIGNNLVY-----KTDIVH-----NQG	848
QY	67 SFLTGPEKLNDRADSRSLWDQGNFLIINKLIKIEDSDTYICEVD-QKEEYQLVFG	125
DB	849 RIV-GDHLMDLAKNR-----QAGTIALERLNLSTANQLTGSGEYNGEYVSKLSFG	900
QY	126 TANSPTHLLOQ-----QSILTLTLESPPG--SSPSVQCRSPRGKNIQGGKTLSTV-SOLEIOD	178
DB	901 -GNLEKPLNDGPNIRINIRTKLSMPVNIITDGDIALAPNDNRSTLQGIKIKTVDSHLNLTG	959
QY	179 SGTWT---CYLONQKVKVERKIDIVLAFOKAS-SIYKKKEGQVEFSPFLATVEKLT	233
DB	960 RANMANIEHMTTELTAQANNFNVDIPSAKXLRFSFNITIKANPKEL-----NLS	1008
QY	224 GSGELMW-----QAEPASSKSMITPDLKNKEVSVR-----VTQDPKL	272
DB	1009 GIVDPMAKRIIDSLPTAEPEVSDEYILNGPHKSKBELIKREPAATKSGMEIRSDLRI	1068
QY	273 QMGKKLPLHLTPALPOYAGSGNLTALBAKTK-----LHDEVNLVVMRAQ-----L	322
DB	1069 NIGKDV-----SLADYGLKTNLDGLSLYKQDKNGNLGULFGQNLTKGRARASFGQDLI	1120
QY	333 QKNTLCEYWGPTSPKLMLSLTKENKEAKVSKREKPVWLNE-----AGMQLISD	374
DB	1121 RKGILSFSGQAOTPELT-----NIEA-----IRNDETMSDKITAGVAVIGIAD	1163
QY	375 SGOVLSENIVKLPTWSPVHPARASALPAPPTGSALPDPOYASALPDPPARASALPALAV	434
DB	1164 SPEVIT-----FSEPSKPDQALSYLITGRSLSSGEGVSGTGVGAA-----	1205
QY	435 ISFLTLGL 443	
DB	1206 ---LIGLGI 1211	
RESULT 38		
NCM2_HUMAN		
ID	NCM2_HUMAN	STANDARD; PRT; 837 AA.
AC	O1534;	
DT	15-JUL-1998 (Rel. 36, Created)	
DT	15-JUL-1998 (Rel. 36, Last sequence update)	
DT	10-OCT-2003 (Rel. 42, Last annotation update)	
DE	Neural cell adhesion molecule 2 precursor (N-CAM 2).	
GN	NCAM2 OR NCAM21.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Brain;	
RX	MEDLINE=97369930; PubMed=9226371;	
RA	Paoletti-Giacobino A., Chen H., Antonarakis S.B.;	
RT	"Cloning of a novel human neural cell adhesion molecule gene (NCAM2)	
RT	that maps to chromosome region 21q21 and is potentially involved in	
RT	Down syndrome."	
RL	Genomics 43:43-51(1997).	
RN	[2]	

RP CARBOHYDRATE-LINKAGE SITES ASN-445 AND ASN-562.
 RX MEDLINE=22660472; PubMed=12754519;
 RA Zhang H., Li X.-J., Martin D.B., Aebersold R.;
 RT "Identification and quantification of N-linked glycoproteins using
 hydrazide chemistry, stable isotope labeling and mass spectrometry";
 RL Nat. Biotechnol. 21:660-666(2003).
 CC -1- FUNCTION: May play important roles in selective fasciculation and
 zone-of-zone projection of the primary olfactory axons.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: Expressed most strongly in adult and fetal
 brain.
 CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U75330; AAB80803.1; -.
 CC Genew; HGNC:7657; NCAM2.
 CC MIM; 602040; -.
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR GO; GO:0005886; C:plasma membrane; TAS.
 DR GO; GO:0007158; P:neuronal cell adhesion; TAS.
 DR InterPro; IPR008957; FN III-like.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_C2.
 DR Pfam; PF00041; FN3; 2.
 DR Pfam; PF00047; Ig; 5.
 DR SMART; SMO0060; FN3; 2.
 DR SMART; SMO0408; IGC2; 5.
 DR PROSITE; PS50835; IG_LIKE; 5.
 DR Cell adhesion; Transmembrane; Glycoprotein; Repeat;
 KW Immunoglobulin domain; Signal.
 KM SIGNAL 1 19
 FT CHAIN 20 837
 FT DOMAIN 20 697
 FT TRANSMEM 698 718
 FT DOMAIN 719 837
 FT DOMAIN 21 108
 FT DOMAIN 113 202
 FT DOMAIN 208 297
 FT DOMAIN 302 396
 FT DOMAIN 401 491
 FT DOMAIN 482 581
 FT DOMAIN 594 678
 FT DISULFID 42 93
 FT DISULFID 136 186
 FT DISULFID 232 281
 FT DISULFID 332 380
 FT DISULFID 422 475
 FT CARBOHYD 177 177
 FT CARBOHYD 219 219
 FT CARBOHYD 309 309
 FT CARBOHYD 406 406
 FT CARBOHYD 419 419
 FT CARBOHYD 445 445
 FT CARBOHYD 474 474
 FT CARBOHYD 562 562
 SQ SEQUENCE 837 AA; 92932 MW; C3D034106C5741C1 CRC64;
 QY Best Match 5.3%; Score 123; DB 1; Length 837;
 QY Query Local Similarity 19.1%; Pred. No. 0.73; Indels 120; Gaps 24;
 Db Matches 92; Conservative 77; Mismatches 192; Indels 120; Gaps 24;
 QY 3 RG-VPEFHLLVLTALLPAAATGKGNVYVGGKGGDTVELCTASOKKSIQHFHKNNSQIKI 61
 Db 193 REEIDPRDIIVTVNPATSMQKSNFNAARGEENTFSCRASGSPRAPISFRNG--KL 250

QY 62 LGNGSFLTKGPKSLNDRADSRSLMDQNFPLIIKNLIKEDSPYICEV-----EDQKE 116
 Db 251 IENNEKYLKG-----SNTBLTVRNIIINSDGPFYVCRANKAGEDKQ 293
 QY 117 E-VOLLVGLTRANSPTHLIQ-----GSLTFLTESPPGSSPSVOC----- 155
 Db 294 AFLQYFV-----QPHIIQIKNETTYENGQ-VTLVCBAEPEPIPEIWKKAVDGFTTE 345
 QY 156 -RSPRG---KNIOGKTLVSQLELODSGTWTCTVLIQ---NOKKVEFKIDIVLAFQ 205
 Db 346 GDKSPDGRIEVGGHSGSSSLHAKVKLSGSGRYCEASRIQGHQKMYLIEVPAP-XFI 404
 QY 206 KASSIVYKKEGEQVFEFSPPLAFVTEKLTGSEELMWQARASSSKSWITFDLKNKEVSYKR 265
 Db 405 SNQITVYSEBNPNISCDV-----KSNPPASIHRRDKL-----VLPKN 445
 QY 266 VTQDPKIQMGKKLPLHLTLPOLPYAGSGNLTALBAKTKLHQEVNLVVMRATOLQKN 325
 Db 446 TTNLKTSTGKMLEIA-PTSDNF-GRVYCT-----ATNHI---GTRPQEX 488
 QY 326 LTCEVWGFTSPKLMISLKLLENKEAKVSKREK-----VWVLNPEAGMQLIS 373
 Db 489 ILALADVPSSEYGVKIIELSQTTAKVS-FNRPDSHGVPINHYQVDYKVASIKWIVRS 547
 QY 374 DSGQVLESNIKVLPTWSTPVHPAPASALPAPPTG--SALPDPQASAL-PDPAPASALPA 430
 Db 548 HGVQYVAVLN-NLEP--NTTYEIRVAVNGKGGQYSKIEIFQILPVREPSPSIHQPS 604
 QY 431 A 431
 Db 605 S 605
 RESULT 39
 ID C166 BRARE STANDARD; PRT; 564 AA.
 AC 090460; 090480;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE C166 antigen homolog precursor (Neuroilin) (DM-GRASP homolog).
 GN CD166.
 OS Brachydanio rerio (zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_Taxid=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94376084; PubMed=8089660;
 RA Kanki J.P., Chang S., Kuwada J.Y.;
 RT "The molecular cloning and characterization of potential chick
 RT DM-GRASP homologs in zebrafish and mouse.";
 RL J. Neurobiol. 25:831-845(1994).
 RN [2]
 RP SEQUENCE OF 398-561 FROM N.A.
 RA MEDLINE=94299040; PubMed=8026643;
 RA Laessing U., Giordano S., Stecher B., Lottspeich F., Stuermer C.A.O.;
 RT "Molecular characterization of fish neuroilin: a growth-associated
 RT cell surface protein and member of the immunoglobulin superfamily in
 RT the fish retinorectal system with similarities to chick protein
 RT DM-GRASP/SC-1/BEN.";
 RL Differentiation 56:21-29(1994).
 CC -1- FUNCTION: CELL ADHESION MOLECULE. INVOLVED IN NEURITE EXTENSION BY
 CC NEURONS VIA HETEROFILIC AND HOMOPHILIC INTERACTIONS.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like V-type domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: L25273; AAA50024.1; -
 CC EMBL: L25057; AAA50048.1; -
 CC HSSP: Q13740; 1K7C.
 DR ZFIN: ZDB-GENE-990415-30; cd166.
 DR InterPro: IPR007110; IG-like.
 DR InterPro: IPR003598; IG_C2.
 DR InterPro: IPR003006; IG_MHC.
 DR Pfam: PF00047; Ig_3.
 DR SMART: SMO0408; IGC2; 1.
 DR PROSITE: PS00835; IG-LIKE; 3.
 DR PROSITE: PS00290; IG_MHC; 1.
 KM Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane;
 KM Repeat; Signal.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 564 CD166 ANTIGEN HOMOLOG.
 FT DOMAIN 25 507 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 508 528 POTENTIAL.
 FT DOMAIN 529 564 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 31 127 IG-LIKE V-TYPE 1.
 FT DOMAIN 131 129 IG-LIKE V-TYPE 1.
 FT DOMAIN 239 323 IG-LIKE C2-TYPE 2.
 FT DOMAIN 340 399 IG-LIKE C2-TYPE 2.
 FT DOMAIN 406 484 IG-LIKE C2-TYPE 3.
 FT DISULFID 38 110 POTENTIAL.
 FT DISULFID 154 217 POTENTIAL.
 FT DISULFID 263 306 POTENTIAL.
 FT CARBOHYD 426 470 POTENTIAL.
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 488 497 LFEEDKPRG -> R (IN REF. 2).
 SQ SEQUENCE 564 AA; 61273 MW; ADAAC7C29A607929 CRC64;
 Query Match 5.3%; Score 122.5; DB 1; Length 564;
 Best Local Similarity 20.6%; Pred. No. 0.477;
 Matches 92; Conservative 62; Mismatches 136; Indels 157; Gaps 21;
 QY 9 HLLVLQALALPA--TOGN--KVLGKKGDTVELTCTASQK--IOFMKNSNQIKI 61
 DB 2 HSVILFGAFIAALFARGSCLPYIGLYGETIEVPCNNGNKPPGLFTMKKA---KD 58
 QY 62 LGNQSFLTKGPKSLNDRADSRSLMDQGNFPLIKLKIEDSDTYICEVEDQKEEVOL 121
 DB 59 DGSFDDLIIKQAK-----DPTVSAMDGKTRV--- 87
 QY 122 VFGFLANSDDTHLQOGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELODST 181
 DB 88 --SIANS-----SLLIKGGSTLDDRV 107
 QY 182 WTCVTLQAKQKVEFKIDIVLAFQKASSIVYKEGEQVFSPLAFVTEKLTGSGELMWQ 241
 DB 108 FTCWVSSVSTNLEEFVEVKV--HKKPSAPVIKNKVKELNG-----KLTQGEQVE 157
 QY 242 AERASSSSWITFDLKNKEVSY-----KRVTDQD-----KLQNGKK----- 277
 DB 158 SANPAADIW-----MKNQALVDDEKTIITSDVTKDVTGLASTSSRLQYARKEVDAS 213
 QY 278 ----LRLHLTLPQALPOYAGSGNLFLAEKTKGKHQEVNLVVMKATQLOK-----NLICE 329
 DB 214 OPTCAKAVTGENV-----STPDPFOAIRYP-----EKVSIQVVSQSPIRGGDDVTLKQ 264
 QY 330 VWGPTSPKLMSTLKENKEAKVSKREKPVWL-----NPEAGWQCLLSDSGVLLLESNIK 385
 DB 265 ADGNRP---ISFNINIKGKVTIVDKOVYITLTGTRADSGVYKCSLLDND--WVESIQI 319
 QY 386 VLPTWSTVPHRASLPA--PPTGSL 410

DB 320 V-----TVSFLDASLTPTGKVL 336
 RESULT 40
 NCAL HUMAN STANDARD; PRT; 848 AA.
 AC P13591; Q15829; Q16180;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Neutral cell adhesion molecule 1, 140 kDa isoform precursor (N-CAM 140)
 DE (NCAM-140) (CD56 antigen).
 GN NCAM1 OR NCAM.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCIT_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94356433; PubMed=8075973;
 RA Saito S., Taniio Y., Tachibana I., Hayashi S., Kishimoto T., Kawase I.,
 RT "Complementary DNA sequence encoding the major neural cell adhesion
 molecule isoform in a human small cell lung cancer cell line."
 RL Lung Cancer 10:307-318(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91250739; PubMed=1710251;
 RA Lanier L.L., Chang C., Azuma M., Ruitenberg J.J., Hemperly J.J.,
 RT "Molecular and functional analysis of human natural killer cell-
 associated neural cell adhesion molecule (N-CAM/CD56)."
 RL J. Immunol. 146:4421-4426(1991).
 RN [3]
 RP SEQUENCE OF 491-848 FROM N.A.
 RX MEDLINE=87301755; PubMed=2887295;
 RA Dickson G., Gower H.J., Barton C.H., Prentice H.M., Elsom V.L.,
 RA Moore S.E., Cox R.D., Quinn C., Putt W., Walsh F.S.;
 RT "Human muscle neural cell adhesion molecule (N-CAM): identification
 of a muscle-specific sequence in the extracellular domain."
 RL Cell 50:1119-1130(1987).
 CC -1- FUNCTION: This protein is a cell adhesion molecule involved in
 CC neuron-neuron adhesion, neurite fasciculation, outgrowth of
 CC neurites, etc.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=N-CAM 140;
 CC IsoId=P13591-1; Sequence=Displayed;
 CC Name=N-CAM 120;
 CC IsoId=P13592-2; Sequence=External;
 CC Name=C; Synonyms=Secreted;
 CC IsoId=P13592-1; Sequence=External;
 CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
 CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD56 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd56.htm".
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).


```

DR InterPro: IPR008957; FN III-like.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003598; Ig_C2.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00047; Ig; 5.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00408; IGC2; 5.
DR PROSITE: PSS0835; IG_LIKE; 5.
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; Repeat; Signal;
KW Transmembrane; Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 848
FT FT 20 848 NEURAL CELL ADHESION MOLECULE 1, 140 kDa
FT FT 20 848 ISOFORM.
FT DOMAIN 20 708 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 709 729 POTENTIAL.
FT DOMAIN 730 730 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 20 111 IG-LIKE C2-TYPE 1.
FT DOMAIN 116 205 IG-LIKE C2-TYPE 2.
FT DOMAIN 212 301 IG-LIKE C2-TYPE 3.
FT DOMAIN 308 403 IG-LIKE C2-TYPE 4.
FT DOMAIN 406 491 IG-LIKE C2-TYPE 5.
FT DOMAIN 518 595 FIBRONECTIN TYPE-III 1.
FT DOMAIN 660 727 FIBRONECTIN TYPE-III 2.
FT DISULFID 41 96 PROBABLE.
FT DISULFID 139 189 PROBABLE.
FT DISULFID 235 287 PROBABLE.
FT DISULFID 329 385 PROBABLE.
FT DISULFID 426 479 PROBABLE.
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 347 347 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 215 215 Q -> R (IN REF. 2).
FT CONFLICT 239 239 L -> F (IN REF. 2).
FT CONFLICT 490 490 QG -> R (IN REF. 3).
FT CONFLICT 599 600 MISSING (IN REF. 3).
FT CONFLICT 720 721 G -> A (IN REF. 3).
FT CONFLICT 811 811 68D2FC0CE6C1C2AD CRC64;
SQ SEQUENCE 848 AA; 93360 MW; 68D2FC0CE6C1C2AD CRC64;

Query Match 5 3%; Score 122.5; DB 1; Length 848;
Beet Local Similarity 20.0%; Pred. No. 0.8;
Matches 61; Conservative 45; Mismatches 142; Indels 57; Gaps 11;

3 RG-VFPRHLLVLAQLALLPAATQGNKVVVGKGGDTVELTCTASOKKSIQFHKNSNQIKI 61
196 REINFPKQIQVAVNVPTTQAGQNVNATNANIGQSTVLVCDAGEPPEPTMSV----- 247
62 LONQGSFLTKGSPSKLNDRAISRRLWDQGNFPIIKNKIEDSDTYICEVE---DQKEE 117
248 -----TKDGEQIEQEDDEDEKYEYFSDSSQLTIKKVKNDEAEYICIAENKAGEQDAT 299
118 VOLLVFG-----LTASDTHLLQGSITLLESPRGSSPSVQCR-----SPRGKNIQGS- 166
300 IHLKVPFKPKITVEVQTMELQMELEBOVTLTCEASGDPITPTWRTSTRNISSEKTLDDGM 359
167 -----KTLVSOLELQDQSGTWTCTVQ--NOKKVEFKIDIVLAFQKASIIYVYKKG 216
360 VVRSHARVSLTLKISQYTDABEYICTASNTIGQDSQSYLEVOYAPKIQGQVAVYTWEG 419
217 EGVE-----FSFPLAFVTEKLTGSGELWQARASSSSKSWITFDLKN-KEVSVKRYTODP 270
420 NQVNITCEVFAYFA-TIS-----WPRDQLLPSSSNYSNIKIYNTPSASYLEVTGPS 470
QY 271 KIQMG 275
DB 471 ENDFG 475

```

```

ANAL DROME
ID _AMAL DROME STANDARD; PRT; 333 AA.
AC P15364; Q9V3A5;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Amalgam protein precursor.
GN AMA OR BG:DS00276.6 OR CG2198.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=89028670; PubMed=3141062;
RT Seeger M.A., Haffley L., Kaufman T.C.;
RT "Characterization of amalgam: a member of the immunoglobulin
RL superfamily from Drosophila."
RL Cell 55:589-600 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX Celniker S.E., Pfeiffer B.D., Knafels J., Martin C.H., Mayeda C.A.,
RA Palazolo M.J.;
RA "Complete sequence of the Antennapedia complex of Drosophila."
RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Bailew R.M., Basu A., Bendale J., Bayraktaroglu L., Beasly E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Klimek B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasero P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Meruluo G., Milbina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Paclé J.M.,
RA Palazolo M., Peltan G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitlska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong P.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely; TISSUE=Embryo;

```

```

RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarion H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, M23561; AAA28367.1; -.
CC EMBL, AE001572; AAD19797.1; -.
CC EMBL, AE003674; AAF54084.1; -.
CC EMBL, AY051911; AAK93335.1; -.
CC PIR, A31923; A31923.
CC FLYBASE; FBgn0000071; Ama.
CC GO; GO:0005886; C:plasma membrane; IDA.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003598; Ig_c2.
CC Pfam; PF00047; Ig_3.
CC SMART; SM00408; IGC2; 2.
CC PROSITE; PS50835; IG_LIKE; 3.
CC Immunoglobulin domain; Glycoprotein; Membrane; GPI-anchor; Signal;
CC Repeat.
CC SIGNAL 1 23 POTENTIAL.
CC CHAIN 24 ? AMALGAM PROTEIN.
CC PROPEP 25 333 REMOVED IN MATURE FORM (POTENTIAL).
CC DOMAIN 25 128 IG-LIKE V-TYPE.
CC DOMAIN 139 223 IG-LIKE C2-TYPE 1.
CC DOMAIN 230 323 IG-LIKE C2-TYPE 2.
CC DISULFID 46 117 PROBABLE.
CC DISULFID 161 208 PROBABLE.
CC DISULFID 251 307 PROBABLE.
CC CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CONFLICT 83 83 O -> K (IN REF. 1).
CC SEQUENCE 333 AA; 36387 MW; F64473DE3DB25F1 CRC64;

Query Match 5.2%; Score 121.5; DB 1; Length 333;
Best Local Similarity 19.4%; Pred. No. 0.28;
Matches 56; Conservative 50; Mismatches 121; Indels 61; Gaps 9;

QY 10 LLLVQLL-----LPATQGNKVVLGKKDPVELTCTASQKKSIOFW-----KSNQ 58
DB 10 LIFCLAIISLDSVLSAPVLSQISKDVASGVDSVEVNCVTEVEGQLSVSMARPSSESDNS 69
QY 59 IKI-----LGNQ--GSFLTKGPKLNDRADSRSLMDQGNFPLIKLKIEDSDTYIC 109
DB 70 VVLSKRNLSLPDQRYNVTYTBGPKT-----GSAIYTRIONIEVSDMGPIEC 117
QY 110 EV-----EDQKEEVL--VFGLTANSDTLLQSGSLTLLTSPGSSPSV----- 153
DB 118 QVLVATEKVTKLSLQIKTPVIAENTPKSTLVTEGQNLDELTCNANGFPKPTISWAREH 177
QY 154 QCRSRGNKIQGGKTLVSQLELDSSGTWTCTVLONQKKEVKIDIVYLAQKASSTYK 213
DB 178 NAWVAGGHLAEPFLIRISVHRMDRGGYCYCAQNGEQPDRRLIRVVEVERPQIAYVRP 237
QY 214 KEGQEVESFPLAFVEKLTGSGELMWQ-----AERASS 248
DB 238 KIAQNVSHSABLECSVQGYRAPTVVMHKNQVGLQSSRIHEVANTASSS 265

```

```

RESULT 42
LAMP HUMAN STANDARD; PRT; 338 AA.
ID LAMP HUMAN
AC Q1343;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Limbic system-associated membrane protein precursor (LSAMP).
GN LSAMP OR LAMP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=96235133; PubMed=8666243.
RX Plimenta A.F., Fischer I., Levitt P.;
RA "cDNA cloning and structural analysis of the human limbic-system-
RT associated membrane protein (LAMP).";
RL Gene 170:189-195(1996).
CC -1- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.
CC CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF
CC MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH
CC OF THE HYPOCAMPAL MOSSY FIBER PROJECTION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- TISSUE SPECIFICITY: Expressed on limbic neurons and fiber tracts
CC as well as in single layers of the superior colliculus, spinal
CC chord and cerebellum.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. IGLON
CC family.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, U41901; AAC50569.1; -.
CC PIR; JC4776; JC4776.
CC Genew; HGNC:6705; LSAMP.
CC MIM; 603241; -.
CC GO; GO:0007399; P:neurogenesis; TAS.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003598; Ig_c2.
CC Pfam; PF00047; Ig_3.
CC SMART; SM00408; IGC2; 2.
CC PROSITE; PS50835; IG_LIKE; 3.
CC Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
CC Repeat; Signal; Lipoprotein.
CC SIGNAL 1 28 POTENTIAL.
CC CHAIN 29 315 LIMBIC SYSTEM-ASSOCIATED MEMBRANE
CC PROPEP 316 338 REMOVED IN MATURE FORM (POTENTIAL).
CC DOMAIN 29 122 IG-LIKE C2-TYPE 1.
CC DOMAIN 132 214 IG-LIKE C2-TYPE 2.
CC DOMAIN 219 304 IG-LIKE C2-TYPE 3.
CC DISULFID 53 111 POTENTIAL.
CC DISULFID 153 197 POTENTIAL.
CC DISULFID 239 290 POTENTIAL.
CC CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC LIPID 315 315 GPI-anchor amidated asparagine
CC SEQUENCE 338 AA; 37308 MW; 03455F286DF5D92F CRC64;
SQ

```

Query Match 5.2%; Score 121.5; DB 1; Length 338;
 Best Local Similarity 20.7%; Pred. No. 0.28; Indels 143; Gaps 22;
 Matches 86; Conservative 63; Mismatches 123

10 LLLVQLALLPAA-----TQGAHVYGGKGDVETLTCTASQKKSIOFHKNSNQKI 61
 14 LVLALLCLLPGLVRSVDPRGTDNITVRGDTAILRCVLEDKNS-KVAMLNRSGLIF 72
 62 LAGNQSFLTKGSPKLNDR--DSRSLMDQGNFPLIIKMLKTEDSDTYICEVEDQK--- 116
 73 AGHD-----KMSLDPRVELEKHSLS--EYSRIQKVADVDEGSYTCSTQTOHEBKT 121
 117 -EVOLLVGG---LANSPTHLLOGOSLTLTSPGSSPSVQCR--SPRGKNIQGR- 168
 122 SOVYLLIVQVPPKISISSDVTVNBSNVTLVCMANGREPEVITWHLPFTGFEFGESEY 181
 169 LSVSOLLELDSSGTWTCTVLQONQKVEFKIDIVLAFQKASSIVYKKEGQVE--PSPFLA 226
 182 LEILGITRQSGKYEC-----KXANVSSADVQKVQVTVNYPT 220
 227 FTVEKLTGSGEL-----WMOAERASSSKSWITFDLKNKE----V 261
 221 ITESK---SNEATTRQASLKCEASAVPAPDEFWYRDDTRINSANGL--EIKSTEGQSL 275
 262 SVKRVTPDPKLOMGKKPLHLPLPOLPYAGSGVLTALAEKTKGKHQGVNLYVMRATQ 321
 276 TVTNVTEE-----HY---GNVTCVANKLVNTN--ASLVFR--- 307
 322 LQKNLTCEVWGPTSPKLM---LSLKENKEAKVSKREKPVVNLNPEAGMOCLLS 373
 388 -----PGSVRGINGSISLAV-----PLWLL--ASLSLCLLS 336

RESULT 43
 KILN RAT STANDARD; PRT; 348 AA.

AC 092038;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Kilon protein precursor (Kindred of Iglon).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-62.
 RX MEDLINE=99175207; PubMed=10075727;
 RA Funatsu N., Miyata S., Kumanooh H., Shigeta M., Hamada K., Endo Y.,
 RA Sokawa Y., Maekawa S.;
 RT "Characterization of a novel rat brain glycosylphosphatidylinositol-
 RT anchored protein (Kilon), a member of the Iglon cell adhesion molecule
 RT family";
 RT J. Biol. Chem. 274:8224-8230 (1999).
 RL J. Biol. Chem. 274:8224-8230 (1999).
 CC -1- FUNCTION: CELL-ADHESION (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- TISSUE SPECIFICITY: Highly expressed in brain.
 CC -1- PTM: Glycosylated.
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. Iglon
 CC family.
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL, AB017139; BAA75649.1; -
 CC DR InterPro, IPR007110; Ig-like.

DR InterPro, IPR003598; Ig_c2.
 DR Pfam, PF00047; Ig_3.
 DR SMART, SM00408; IGC2_2.
 DR PROSITE, PS50835; IG-LIKE_3.
 DR Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
 KW Repeat; Signal.
 FT SIGNAL 1 31
 FT CHAIN 32 31
 FT PROPER 32 31
 FT DOMAIN 32 128
 FT DOMAIN 133 215
 FT DOMAIN 219 307
 FT DISULFID 54 112
 FT DISULFID 154 197
 FT DISULFID 239 291
 FT CARBOHYD 67 67
 FT CARBOHYD 149 149
 FT CARBOHYD 269 269
 FT CARBOHYD 280 280
 FT CARBOHYD 288 288
 FT CARBOHYD 301 301
 SQ SEQUENCE 348 AA; 37858 MW; 37E90D1C7D2AACA6 CRC64;

Query Match 5.2%; Score 121; DB 1; Length 348;
 Best Local Similarity 23.1%; Pred. No. 0.32;
 Matches 74; Conservative 51; Mismatches 116; Indels 80; Gaps 19;

20 PAATQGNKVYGVKGGDTVELTCTASQKKSIOFHKNSNQIKILNQSFLTKGP---SK 75
 35 PMAVVDNMLV--RKGDPTVLRCLYLEDGAS-KGAMLNRSGLIFAG--GDKMSVDPRVSI 89
 76 LNDADSRSLMDQGNFPLIIKMLKTEDSDTYICEVEDQKQEVOLVGLTA----- 127
 90 LNKR-----DVSLOQNVDTVDGPTGTSVQTOPTPTMOVH-LTVQVPPKIYD 137
 128 -NSDTHLLOGOSLTLTSPGSSPSVQCR--SPRGKNIQGGKTLVSVQLELDSSGTWTC 184
 138 ISNMTINEGTVNLTCLATGKPEPAISWRIHISAPKFEQVLYDIGTRDQAGEVYC 197
 185 TVLONQKVEF---KIDIVV---LAFQASST-----VYKKEGQVEFSPPLATV 229
 198 SA---ENNVSPDVQKRVVNVFAPPTIOEIKSGVTPGSLIRCEGAVP--PPAPE- 250
 220 EKLQSSGLMWQARRA---SSSKSWITFDLKNKEV-SVKRVYTD-----PKL-QMG 275
 251 -----WYKGEGRFLNGQGGITIQNFSTRSLITVNTVQEHGNTVCAANKLGITT 301

RESULT 44
 SCPI RAT STANDARD; PRT; 997 AA.

AC 003410;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 15-JUN-1998 (Rel. 36, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Synaptonemal complex protein 1 (SCP-1 protein).
 GN SCPI OR SCPI.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93099684; PubMed=1464329;
 RX Meuwissen R.L.J., Offenberg H.H., Dietrich A.J., Riesewijk A.,
 RA Ieriel M., Heyting C.;
 RA "A coiled-coil related protein specific for synapsed regions of
 RT meiotic prophase chromosomes";

```

RL EMBO J. 11:5091-5100(1992).
CC -1- FUNCTION: Major component of the transverse filaments of
CC synaptonemal complexes (SCs), formed between homologous
CC chromosomes during meiotic prophase.
CC -1- SUBCELLULAR LOCATION: Nuclear. In tripartite segments of
CC synaptonemal complexes, between lateral elements in the nucleus.
CC Found only where the chromosome cores are synapsed. Its N-terminus
CC is found towards the centre of the synaptonemal complex while the
CC C-terminus extends well into the lateral domain of the
CC synaptonemal complex.
CC -1- TISSUE SPECIFICITY: Testis.
CC -1- DEVELOPMENTAL STAGE: Expressed exclusively in meiotic prophase
CC cells.
CC -1- DOMAIN: Consists of an alpha-helical stretch of 700 AA residues,
CC flanked by N- and C-terminal globular domains. The C-terminal
CC domain has DNA-binding capacity.
CC -1- CAUTION: This is a conceptual translation; a frameshift was
CC corrected in position 6 to maximize the similarity with the
CC other species SYCP1 sequences.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X67805; CA48006.1; ALT_FRAME.
DR InterPro: IPR008827; SCP-1.
DR Pfam: PF05483; SCP-1;
KM Nuclear protein; Meiosis; Cell division; Phosphorylation;
FM DNA-binding; Coiled coil.
FT DOMAIN 108 819 COILED COIL (POTENTIAL).
FT DOMAIN 118 121 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 701 704 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 902 905 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 982 990 ARG/LYS-RICH (BASIC).
SQ SEQUENCE 997 AA; 116511 MW; 229D59823FD684BE CRC64;

Query Match 5.2%; Score 121; DB 1; Length 997;
Best Local Similarity 19.8%; Pred. No. 1.2; Indels 118; Gaps 24;
Matches 96; Conservative 95; Mismatches 175;

QY 9 HLLVLTQ-----ALPAAQGNKVKVGGKDPVELTCTASQKSIQF-----WKNQNOIK 60
DB 374 HSLVTELTAKTCTLEELIRTEQGLENNEDQKIMELQKSELEBMKFKNNKVE 433
QY 61 ILNGSGFLTKGPKSLNDR-----ADSRRLMDQGNFPLIITKNKIED----- 103
DB 434 -LEELKTLIAEDQKLLDEKKQVEKLAELQKGEQELTFLQTRKEKIHDELVQVTVTKTS 492
QY 104 SDTYCEVEDQKEVQ-----LVFGLTANSDPHILQGGSLT-----LTLESPPGSSPSVQC 155
DB 493 EBHYKQVEEMTELEKEKLTNLTANSDDLLENKKLVODASDMVELLKKHEDIIINC 552
QY 156 RSPRKNIOGKTLISVSOLEQDS-GTWTCTVLQKQKVEFKIDIVLAFQKASSIVY-- 212
DB 553 KKQERMLKQIETLEKEMNLDELSEVRKEFIQGGDEVKCKLD--KSEENASIEVEV 609
QY 213 -KKEGE-----QVEFSPLAFVTEKLTGSGELMWQERASSSKSWTFDIK- 257
DB 610 LKKEKQMKILENKKNNLKQIENK--SKNIEELHQENKA-LKKKSSANKKQLNAVEIK 665
QY 258 -----NKEYSVRVYVODPRLQMGKKLPLHLTLPOALPOYAGSGN 296
DB 666 NKLIELASTQKPEEMINNTYQKELEIKKISE-KL-LGEVEKAKATYDEAVKUL--QKE 720
QY 297 LTLALEATGKL-----HOEVALVVMRATQ-LQKTLCEVWGPTSPKMLSTIKLEN 346
DB 721 IDLRQCHIAEVALMEKHQYDKIYVERDSEGLYKNRQF--QSSAYALETELISN 777
QY 347 -----KEAKVSKREKRVVVLNPEAGMOCILSDSGVLLBSNIRKVL--TW----- 390

```

```

DB 778 IRNELVSLKQLVEKEKEKELKMEQEN---TAILTDKKDKIKQASLSEPSATSWRFD 834
QY 391 -STP 393
DB 835 KTTIP 838

RESULT 45
PIGR_RABIT STANDARD; PRT; 773 AA.
AC P01832;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Polymeric-immunoglobulin receptor precursor (Poly-Ig receptor) (PIGR)
DE [Contains: Secretory component].
GN PIGR.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RX MEDLINE=84142246; PubMed=6322002;
RA Mostov K.E., Friedlander M., Blobel G.;
RT "The receptor for transepithelial transport of IGA and IGM contains
RT multiple immunoglobulin-like domains.",
RL Nature 308:37-43(1984).
RN [2]
RP SEQUENCE OF 87-114 AND 410-428.
RX MEDLINE=88228032; PubMed=1131339;
RA Frutiger S., Hughes G.J., Hanly W.C., Jaton J.-C.;
RT "Rabbit secretory components of different allotypes vary in their
RT carbohydrate content and their sites of N-linked glycosylation.";
RL J. Biol. Chem. 263:8120-8125(1988).
CC -1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE
CC TRANSMEMBRANE SEGMENT.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Also secreted.
CC -1- POLYMORPHISM: The sequence shown is that of allotype T62.
CC -1- SIMILARITY: Contains 5 immunoglobulin-like V-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X00412; CA425118.1; -.
DR PIR: A02111; ORRBG.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003599; IG.
DR Pfam: PF00047; Ig; 5.
DR SMART: SM00409; Ig; 5.
DR PROSITE: PS50835; IG LIKE, 3.
KM Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein; Signal;
KM Polymorphism.
FT SIGNAL 1 18
FT CHAIN 19 773 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.
FT CHAIN 19 615 SECRETORY COMPONENT.
FT DOMAIN 19 647 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 648 670 POTENTIAL.
FT DOMAIN 671 773 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 25 131 IG-LIKE V-TYPE 1.
FT DOMAIN 138 232 IG-LIKE V-TYPE 2.
FT DOMAIN 233 340 IG-LIKE V-TYPE 3.
FT DOMAIN 352 455 IG-LIKE V-TYPE 4.

```

```

FT DOMAIN 461 557 IG-LIKE V-TYPE 5.
FT DISULFID 46 115 POTENTIAL.
FT DISULFID 155 225 POTENTIAL.
FT DISULFID 260 324 POTENTIAL.
FT DISULFID 369 438 POTENTIAL.
FT DISULFID 478 538 POTENTIAL.
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (IN ALLOTYPES T61).
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (IN ALLOTYPES T62).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (IN ALLOTYPES T63).
FT VARIANT 88 88 K -> N (IN ALLOTYPES T61).
FT VARIANT 94 94 D -> E (IN ALLOTYPES T61).
FT VARIANT 101 108 TYDOLTON -> YINRISLOS (IN ALLOTYPES T61).
FT VARIANT 110 110 S -> T (IN ALLOTYPES T63).
SQ SEQUENCE 773 AA; 8386 MW; DF2C44D2F1193C65 CRC64;

Query Match 5.2%; Score 120.5; DB 1; Length 773;
Best Local Similarity 20.3%; Pred. No. 0.97;
Matches 108; Conservative 75; Mismatches 199; Indels 149; Gaps 26;

QY 5 VFFRHL-----VLQALLPAATQGN-----KVLGKKGTVELTCTASQKKSIO 50
DB 210 VIKHLQNLNDAGQYVQSGSDPTAEQNDRLRLTGLLYGNIGSGVFECALDSEBDNA 269
QY 51 FPMKSNQK-----ILNGSGFLTKGPKSLNDRASRLM---DQGNPLIKLIKIED 103
DB 270 V--ASLRVGRGNVVDGGTI-----DPAFGRLLFTKAENGHSVIAIGRKED 318
QY 104 SDTYICEVDQKEVOLVFGLTANSDDLQ---GQSILTLTLESP-----PGSSPVQC 155
DB 319 TGNVLCGVQSN-----GQSGDPTQLRQLFVNEIDVGRSPVLKGFQGSVTRIC 369
QY 156 -RSPP-----GKNI-----QGKITSV--SQLE 175
DB 370 PYNPKRSDSHLQLYMEGSGQTHLLVDSGEGVLQDYTRLALFEPFGGTSSVVLNQLT 429
QY 176 LDPSGTCTVQLONOKVEFKIDIVLAFQKASSI--VYKKEGQVEFS--PPLAFVYK 231
DB 430 ADEGFGYWC-VSDDDESILTSYKQLQVDESPPTIDKFAVQGEPEITCHFPCKY---- 484
QY 232 LTGSEGLMWQAERASSSKSMITF-DLKNKEVSKRVYTOPDKLQMGKULPLHLTPQALPQ 290
DB 485 -----FSEKWKYCKMNDHGCDLPTKLSSGDLVKNNNLVLTLTDSVSD 531
QY 291 YG---SGVLTALAEKTKLHGVNLVVRATLOLKNLTCEWGPSTSKMLSLKLEK 347
DB 532 DEGWYWC-----AKDGHFEFEVAAVVELTEPAKVAVEPAKVPVDPKAAAPAPAEK 584
QY 348 EAKVSKREKRVVVLNPEAGMQLDSDQVLLSENIKVLPTWSPVHPRASALDAPPTG 407
DB 585 ---AKARCPV---PRRQWYPL-----SRKLRT-SCP-EPRLIAEEVAVOS 622
QY 408 SALPPQTASALPDPPASALPALAVISFLGLGI-----GVACYLART 453
DB 623 ADPPASGASVYDASASGSGSAKYLITVPLGLVLAAGMAVAIAAR 673

RESULT 46
NCAL CHICK STANDARD; PRT; 1091 AA.
AC P13590; Q90918; Q90919;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 1, 180 kDa isoform precursor (N-CAM
DE 180).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 1-175 FROM N.A., AND PARTIAL SEQUENCE.

```

```

RX MEDLINE=87206190; PubMed=3576199;
RA Cunningham B.A., Hemperly J.J., Murray B.A., Prediger E.A.,
RA Brackenbury R., Edelman G.M.;
RT "Neural cell adhesion molecule: structure, immunoglobulin-like
RT domain, cell surface modulation, and alternative RNA splicing.";
RL Science 236:799-806(1987).
[2]
RX SEQUENCE OF 128-1091 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=86206089; PubMed=3458261;
RA Hemperly J.J., Murray B.A., Edelman G.M., Cunningham B.A.;
RT "Sequence of a cDNA clone encoding the polysialic acid-rich and
RT cytoplasmic domains of the neural cell adhesion molecule N-CAM.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3037-3041(1986).
[3]
RX SEQUENCE FROM N.A. (ISOFORM D).
RX MEDLINE=87092340; PubMed=3467341;
RA Hemperly J.J., Edelman G.M., Cunningham B.A.;
RT "cDNA clones of the neural cell adhesion molecule (N-CAM) lacking a
RT membrane-spanning region consistent with evidence for membrane
RT attachment via a phosphatidylinositol intermediate.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9822-9826(1986).
[4]
RX SEQUENCE OF 810-1069 FROM N.A.
RX MEDLINE=87033934; PubMed=3771645;
RA Murray B.A., Owens G.C., Prediger E.A., Crossin K.L.,
RA Cunningham B.A., Edelman G.M.;
RT "Cell surface modulation of the neural cell adhesion molecule
RT resulting from alternative mRNA splicing in a tissue-specific
RT developmental sequence.";
RL J. Cell Biol. 103:1431-1439(1986).
[5]
RX SEQUENCE OF 1-17 FROM N.A.
RX MEDLINE=93122797; PubMed=1478668;
RA Colwell G., Li B., Forrest D., Brackenbury R.;
RT "Conserved regulatory elements in the promoter region of the N-CAM
RT gene.";
RL Genomics 14:875-882(1992).
[6]
RX SEQUENCE OF 1-17 FROM N.A.
RX STRAIN=white leghorn; TISSUE=erythrocyte;
RA Saecker M., Covault J.;
RT Submitted (AUG-1993) to the EMBL/Genbank/DBJ databases.
CC -! FUNCTION: This protein is a cell adhesion molecule involved in
CC neuron-neuron adhesion, neurite fasciculation, outgrowth of
CC neurites, etc.
CC -! ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=A; Synonyms=N-CAM 180;
CC IsoId=P13590-1; Sequence=Displayed;
CC Name=B; Synonyms=N-CAM 140;
CC IsoId=P13590-2; Sequence=VSP_002585;
CC Name=C;
CC IsoId=P13590-3; Sequence=VSP_002586;
CC Name=D;
CC IsoId=P13590-4; Sequence=VSP_002583, VSP_002584;
CC -! SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC -! SIMILARITY: Contains 2 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15861; AAB59958.1; JOINED.
DR EMBL; M15860; AAB59958.1; JOINED.
DR EMBL; M15922; AAB59958.1; JOINED.
DR EMBL; M15923; AAB59958.1; JOINED.
DR EMBL; M15924; AAB59958.1; JOINED.
DR EMBL; M21178; AAB59958.1; JOINED.
DR EMBL; M21179; AAB59958.1; JOINED.

```

FT	DOMAIN	152	156	HEPARIN-BINDING (POTENTIAL) .
FT	DOMAIN	161	165	HEPARIN-BINDING (POTENTIAL) .
FT	DISULFID	41	96	BY SIMILARITY .
FT	DISULFID	139	189	BY SIMILARITY .
FT	DISULFID	235	287	BY SIMILARITY .
FT	DISULFID	329	385	BY SIMILARITY .
FT	DISULFID	426	479	BY SIMILARITY .
FT	CARBOHYD	222	222	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	315	315	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	347	347	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	423	423	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	449	449	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	CARBOHYD	478	478	N-LINKED (GLCNAC . . .) (POTENTIAL) .
FT	VARSPLIC	702	726	STSPSGGCTAATVGLIVLIVFILL -> TLGSPGSSGFV
FT				SLLSAATLLLC (in isoform D) .
FT	VARSPLIC	727	1091	/FTID=VSP_002583 .
FT				Missing (in isoform D) .
FT	VARSPLIC	771	809	/FTID=VSP_002584 .
FT				Missing (in isoform B) .
FT	VARSPLIC	810	1070	/FTID=VSP_002585 .
FT				Missing (in isoform C) .
FT				/FTID=VSP_002586
SQ	SEQUENCE	1091 AA; 117415 MM; B151367002DF8BE0 CRC64;		
Query Match		5.2%; Score 120.5; DB 1; Length 1091;		
Best Local Similarity		20.0%; Pred. No. 1.5;		
Matches		84; Conservative 52; Mismatches 157; Indels 127; Gaps 16		
OY	32	KKGDTVELTCTAAGKKSIGFHWKNSNQIKILGNSFLTKGSKLNDRAISRSLMDGN	91	
DB	130	KEGDVAIVCVVSSLPETIIWKKHGRVMLKKDVRIV-----LSN	171	
OY	92	PFLIIKNIKIDSDTYICE-----VEDOKEEVQLV-----FGLTANSDTL	133	
DB	172	NYLQIRGIKKTDEGTYRCERGIILARGEINFPDIOQIVAVVPSVRAROSTMTANT-----	226	
OY	134	LOGOSLTITLESPOSSPSPVOCRSBRKNIQ-----GGKTLSVSOLEIJDST	181	
DB	227	LSOSVTILACADGFPETPTMTW--TKDGEPIEODNEEKYSFVNDSESLIKKYVDSDAE	284	
OY	182	WTCVVLNQKQKVEFIDIVLAFOKASIVYKKEGEJVFSPPLAFYTEKLTGSGE----	237	
DB	285	YITLAEKAKAGQDATHLKVAKPK--ITYVENTAMELEQITLTC--ASGDPIPS	338	
OY	238	LMWQAEARASSKSWITFDLKNKEVSVKRVTDPKLQMGKCLPLHLTPQALPOYAGSGL	297	
DB	339	ITWKSTSTNISNEKTLIDGRIVVSHARVSS-----LTLKEI--QYTDAGEY	383	
OY	298	TLALEAKTKGLHGEVNLVVMATQL-----QKULTCEVWGFPSPKLTMSIKLE	345	
DB	384	VCTASNTIGQSOAMYLEVOYAPKLQGFVAAYVTEGNOVNTICEVF--AYPSAVIS----	437	
OY	346	NKEAKVSKREKPVWLVLEAGMOCILSDSCQVLLNESNIKVLPTWSTVHPAPASALPAP	405	
DB	438	-----WFRD-----GQLLPSSVYSNIKIYNT---PSASTLEVTIP	468	
RESULT 47				
APLP MANSE STANDARD; PRT; 3305 AA.				
ID	APLP MANSE	STANDARD;	PRT;	3305 AA.
AC	025450;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Apolipophorin precursor [Contains: Apolipophorin-I (APOLP-1);			
OS	Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).			
OC	Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Spingioidea;			
OC	Springidae; Sphinginae; Manduca.			
NCBI	_taxid=7130;			
RN	[1]			
RP	SEQUENCE FROM N.A.			

```

RC TISSUE=Fat body;
RX MEDLINE=97166603; PubMed=9014323;
RA Sundermeyer K., Hendricke J.K., Prasad S.V., Wells M.A.;
RT "The precursor protein of the structural apolipoproteins of
RT lipophorin: cDNA and deduced amino acid sequence.";
RT Insect Biochem. Mol. Biol. 26:735-738(1996).
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; SECRETED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, U57651; AAB53254.1; -.
DR PIR, T18358;
DR InterPro: IPR001747; Lipid_transprt_N.
DR InterPro: IPR001846; VWF_D.
DR Pfam, PF01347; Vitellogenin_N; 1.
DR Pfam, PF00094; vwd; 1.
DR SMART, SM00638; LPD_N; 1.
DR SMART, SM00216; VWD; 1.
KW Signal.
FT SIGNAL 1 23
FT CHAIN 24 ?
FT CHAIN ? 3305
FT SEQUENCE 3305 AA; 366941 MW; B27D611410285FD7 CRC64;
SQ
Query Match 5.2%; Score 120.5; DB 1; Length 3305;
Best Local Similarity 22.5%; Pred. No. 6.4; Mismatches 176; Indels 93; Gaps 21;
Matches 97; Conservative 66;
26 NKRYLGGKGDVLTCTASQKKS-----IQFMKNSNQIKIGNGSFLTKGSPKSLNDRA 80
1852 NSVVVDADGRVYKIDSSIVLSKAHPVLDIQHSPSPDKIRLYLQGSLSSTQGLEVKV 1911
81 DSRRLMDQGNFPLIKIKIEDSDTYICEVEDQKEVQLVFGLTANSDTLLQGQSIT 140
1912 DN-----INDICLD-----AVSEANVQKDN---AFKVANAKELGWNKYGID 1951
141 LTLSPGSSPVSOCRSPPR-GKNIOGKTLVSOLELDS--GTWTCVTLNQKKEPK 196
1952 IS-SKDSGSGKRLFHATNDKNVLSGSTSFISKQGGTTIEGSGSVKVEEQSANPK 2010
197 IDIVVIAFOKASSIYVKKEGEO-VESFPPLAFTVEKLTGSGELMWQAERASSSSKSWIT-F 254
2011 Y-----IRYVFTDSNKEKGEVEFFVVAL-----GERSYVAESRVTNY 2046
255 DLKAKKESVAVKVTQDPKLOMGKLLPLHLTLPOALPOYAGSGMLTLALEAKTGKLEHVN 314
2047 EYKNSYVVCSEKKQCAHAEIQSK--IDMSTPGMIVNVINAG-LDL---RKLGV----- 2093
315 VVMRAATOLQKNTCEVWGTPSPKMLSLKLENKEKVSREKRVNVLN-PEAGMMQCLLS 373
2094 ----APRGLQWRDEVSDDRRPRFTLDLHT-NKE-----DRKHYLAAVTPPEGHY----A 2140
374 DSGQVVLBSNIVKLPFTWSTPVHPRASALPAPRTGSLPDP-----DOTASALPDP 425
2141 SGVTVRLPSPRWALE--YTLTHPTSGDLPFPFKGEACLDLDGNRPGHKTSAFLVDYNS 2198
426 SALPALAVISF 437
2199 GSEDKAVVAEIGF 2210

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Platelet endothelial cell adhesion molecule precursor (PECAM-1)
DE (CD31 antigen).
GN PECAM1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxId=99823;
RN [1]
RP SEQUENCE FROM N.A.
RA Nasu K.;
RL FUNCTION: (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE EXPRESSED ON
CC PLATELETS AND AT ENDOTHELIAL CELL INTERCELLULAR JUNCTIONS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X98505; CA67129.1; -.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003599; IG.
DR Pfam, PF00047; Ig; 4.
DR SMART, SM00409; Ig; 3.
DR PROSITE, PS50835; IG_LIKE; 4.
KW Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
KW Repeat; Signal.
FT SIGNAL 1 27
FT CHAIN 28 740
FT DOMAIN 28 602
FT TRANSMEM 603 621
FT DOMAIN 622 740
FT DOMAIN 35 126
FT DOMAIN 145 223
FT DOMAIN 236 315
FT DOMAIN 328 404
FT DOMAIN 425 494
FT DOMAIN 500 592
FT DISULFID 57 109
FT DISULFID 152 206
FT DISULFID 256 304
FT DISULFID 347 387
FT DISULFID 432 477
FT DISULFID 524 573
FT CARBOHYD 52 52
FT CARBOHYD 84 84
FT CARBOHYD 284 284
FT CARBOHYD 301 301
FT CARBOHYD 320 320
FT CARBOHYD 357 357
FT CARBOHYD 372 372
FT CARBOHYD 436 436
FT CARBOHYD 456 456
FT CARBOHYD 552 552
SQ SEQUENCE 740 AA; 82378 MW; F312DC624B4A217 CRC64;
Query Match 5.1%; Score 118.5; DB 1; Length 740;
Best Local Similarity 21.6%; Pred. No. 1.3;
Matches 75; Conservative 58; Mismatches 116; Indels 99; Gaps 19;
101 IEEDDTYICEVEDQK-BEVQLVFGLT-----ANSDTLLQGQSITLTLSPGSSP 151
296 VEHNSNTCKVEARISKVSIMVITELFSPRLKKSATRLDGGESIRLMC-SIPGAP 354
152 SVQKSPRGKNIQGGKTLVSQLEL-----QDSGTWTC-----YLNQKQKVEFKDI 200

```

```

Db 355 EAMF-----TIQKGMMLDQDNLTKVASERDSGTTCVAGIGKVKRSNEVOJAVCEM 408
Oy 201 V---LAFQKASIIYKKEGEVFEFS-----FPLAFTVEKLTGSGELMWOAEKRSASSKS 250
Db 409 LSKPSIFDSSGSEVKK--GQITIEVSCGSIINGTSPISTY--QLIKSSDLL--ASQNVSSNEP 462
Oy 251 WITFDLKNKEVSVKRVTDPKLQMGKULPLHLTLPLQALPOYAGSGNLTLEAKTGKLIHQ 310
Db 463 AVFKONPKRVDEYQCIADN-----CHSHAGMPSKVLRVKVIAPVE 502
Oy 311 EVNLVVMATQLOQRK---LTCEV---WGPTSPKMLSLKLNKAKVSKREKP----- 357
Db 503 EVKLISLISEEVSGQALVLQCSYKSGSGPTTKFY-----KE---KENKPFHQVTL 551
Oy 358 -----VM-----VLNPEAGMOCCLSD---SGOVLESNIKVLPTWSTP 393
Db 552 NDTQAIWKPKRASKQEGQYICLASNRATPSKNFLQSIILAVRYLAP 599

RESULT 49
LAMP RAT STANDARD; PRT; 338 AA.
AC Q62813;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Limbic system-associated membrane protein precursor (LSAMP).
GN LSAMP OR LAMP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N. A., AND SEQUENCE OF 29-49.
RC TISSUE=hippocampus;
RX MEDLINE=95374785; PubMed=7646886;
RA Pimenta A.F., Zhukareva V., Barbe M.F., Relinosa B.S., Grimley C.,
RA Henzel W., Fischer I., Levitt P.;
RA "The limbic system-associated membrane protein is an Ig superfamily
RT member that mediates selective neuronal growth and axon targeting.";
RT Neuron 15:287-297(1995).
RL
CC -I- FUNCTION: MEDIATES SELECTIVE NEURONAL GROWTH AND AXON TARGETING.
CC CONTRIBUTES TO THE GUIDANCE OF DEVELOPING AXONS AND REMODELING OF
CC MATURE CIRCUITS IN THE LIMBIC SYSTEM. ESSENTIAL FOR NORMAL GROWTH
CC OF THE HYPOCAMPAL MOSSY FIBER PROJECTION.
CC -I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -I- TISSUE SPECIFICITY: Expressed mostly by neurons comprising limbic-
CC associated cortical and subcortical regions that function in
CC cognition, emotion, memory, and learning.
CC -I- DEVELOPMENTAL STAGE: FIRST DETECTED AT E15-16, AT STAGE E20 IT IS
CC DETECTED IN PRESUMPTIVE CORTEX, MEDIAL LIMBIC AREAS OF THE
CC THALAMUS AND HYPOTHALAMUS. IN THE ADULT, IT IS FOUND IN
CC HYPOTHALAMUS, PERIRHINAL CORTEX, AMYGDALA AND MEDIAL THALAMIC
CC REGION.
CC -I- SIMILARITY: Belongs to the immunoglobulin superfamily. IGLON
CC family.
CC -I- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U13554; AA86120.1; .
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003598; Ig_C2.
CC Pfam; PF00047; Ig_3.
CC SMART; SM00408; IGC2; 2.
CC PROSITE; PS50835; Ig_LIKE; 3.
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;

```

```

KW Repeat; signal; Lipoprotein.
FT SIGNAL 1 28
FT CHAIN 29 315
FT PROPEP 316 338
FT DOMAIN 29 122
FT DOMAIN 132 214
FT DOMAIN 219 304
FT DISULFID 53 111
FT DISULFID 153 197
FT DISULFID 239 290
FT CARBOHYD 40 40
FT CARBOHYD 66 66
FT CARBOHYD 136 136
FT CARBOHYD 148 148
FT CARBOHYD 279 279
FT CARBOHYD 287 287
FT CARBOHYD 300 300
FT CARBOHYD 315 315
FT LIPID 315 315
SQ SEQUENCE 338 AA; 37324 MW; 0B76AFDD68A39BB6 CRC64;

Query Match 5.1%; Score 117.5; DB 1; Length 338;
Best local similarity 20.5%; Pred. No. 0.53;
Matches 85; Conservative 64; Mismatches 123; Indels 143; Gaps 22;

Oy 10 LLVLVQLALPLPA-----TQGNKVLGKKGDPVETLTCTAGQKSIQPHMKNSNQIKI 61
Db 14 LVLRRLCLPLPTGLPVRSDVFNRTGDNITVYQGGTALRLCVEDKNS-KVAMLRSGIIF 72
Oy 62 LGNGSFLTKGPKSLNDRA--DSRRSLMDQGNPLIINKKIEDSDTYICEVEDQKE--- 116
Db 73 AGHD-----KMSLDPRVLEKRNAL-----EVSRLQKVDYDEGTSVQGHPKT 121
Oy 117 -EYQLVFG---LTNSDTHLIQGSGLTTLSPRESSSVQCR--SPGKNIQSGKT- 168
Db 122 SOVYLLIQVPPKISNISSDVTNVEGNSVTLVCNANGRPREVITWRHLTPLGREFEGEEY 181
Oy 169 LVSQLELDQSGTWTCTVQLQNKVKFEKIDIVLAFQKASIIYKKEGEVFE--FSPLA 226
Db 182 LEIIGITREDSGRKEC-----KAAVEVSSADVKQKVTYVNPPT 220
Oy 227 FVEKLTGSGEL-----WQAEKRSASSKSWITPLKNE---V 261
Db 221 ITSEK---SNEATTGQASLKCEASAVPAPDFEYRDRDTRINSANG--EIKSTEGGSSL 275
Oy 262 SVKRVTDPKLQMGKULPLHLTLPLQALPOYAGSGNLTLEAKTGKHQEVNLVVMRATQ 321
Db 276 TVTNVTEE-----HY---GNVTCVAANKLGVTN--ASLVLFRR--- 307
Oy 322 LQKNLTGEVWGPTSPKLM---LSIKLENKAKVSKREKPVWVNLPEAGMOCCLIS 373
Db 308 -----PGSVGINGSISLAV-----PLMLL--ASLFLCLLS 336

RESULT 50
NCAL RAT STANDARD; PRT; 858 AA.
AC P13596;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell1 adhesion molecule 1, 140 kDa isoform precursor (N-CAM 140)
DE (NCAM-140).
GN NCAM1 OR NCAM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N. A.
RC TISSUE=Brain;

```


[illegible]

FT	CABOHYD	489	489	N-LINKED (GLCNAC..)	(POTENTIAL).
SO	SEQUENCE	858 AA;	94658 MM;	EALIA064A4EA05076	CRC64;
Qy	Query Match		5.1%;	Score 117.5;	DB 1; Length 858;
	Best Local Similarity		19.1%;	Pred. No. 1.8;	
	Matches		79;	Conservative 52;	Mismatches 143; Indels 139; Gaps 17;
Qy	3	RG-VPEPHLLIVLTALLPAATQGNKVVLAGKGGTVELTCTASQKSKIQTFW-KNSNLIK	60		
Db	196	RGEINFDIOIVAVNPPTVOAROSIVATNMLGGSVTLVCDADGPEFTMSWTIDGEP	255		
Qy	61	ILGNQGSFLTGPESKLDNRADRSRLMDQGNFPLIKNLKIEDSDTYICEV----	DOKE 116		
Db	256	-----NEEBDEKHITSDDSSSELTIRVNDKNDAEYCIAMENKAGEODA	299		
Qy	117	EVQLVFG---LTNSDTHLQGSITLTLTLESPGSSPSVQCRSPKNIQGGKTLVS	172		
Db	300	SIHKVFAKPKITVENQTMELBEQVTLTCEASGDPIPSITWMT-----	STR 347		
Qy	173	QLEHDSGTWTCTVLQOKKVEFKIDIVLAFOQASSTIVYKKEGQVEFSPPLAFYVKL	232		
Db	348	NISSEKASWT---REPKEITLDGHVAVSHARVSSITLK-----SIOY-----	388		
Qy	233	TGSEELWQMERASSKSWITFDLKNKEVSKRTQDPKIQMGKKLPMLHLTLPQALPQYA	292		
Db	389	TDAGEYICTASNTIGDS-----QSMYLEVOYAPKIQG-----PVAV--YT	427		
Qy	293	GSGLITLALAEKTKLHQEVNLVVMRATQLOKNTLCEWGCPTSPKMLSLTENKEAKVS	352		
Db	428	WEGN-----QVNTICEVFAAPS-----ATIS	448		
Qy	353	KREKPVVNLPEAGMOCCLSDSGVILLESNIKVLPTWSTPVPHPASALPAP	405		
Db	449	-----WFRD-----COLPSSNYVSNIKIYNT---PSASYLEVTP	479		
RESULT 51					
CAM1	BRARE	STANDARD;	PRT;	1197 AA.	
AC	090478;				
DT	10-OCT-2003 (Rel. 42, Created)				
DT	10-OCT-2003 (Rel. 42, Last sequence update)				
DT	15-MAR-2004 (Rel. 43, Last annotation update)				
DE	Neuronal cell adhesion molecule Li.1 (N-CAM Li.1) (Fragment).				
GN	NADL1.1.				
OS	Brachydanio rerio (Zebrafish) (Danio rerio).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;				
OC	Cyprinidae; Danio.				
OX	NCBI_TaxID=7955;				
RN	[1]				
RP	SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.				
RC	TISSUE=Embryo;				
EX	MEDLINE=96155762; PubMed=8568941;				
RA	Tongigiorgi E., Bernhard R.R., Schachner M.;				
RT	"Zebrafish neurons express two Li-related molecules during early				
RT	axogenesis.";				
RU	J. Neurosci. Res. 42:547-561(1995).				
CC	-I- FUNCTION: Cell adhesion molecule with an important role in the				
CC	development of the nervous system. Involved in neuron-neuron				
CC	adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds				
CC	to axonin on neurons (By similarity).				
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.				
CC	-I- TISSUE SPECIFICITY: Expressed in postmitotic neurons in 16-36				
CC	hour embryos; including those in the brain, cranial ganglia and				
CC	otic and olfactory placodes, and in all classes of spinal				
CC	neurons.				
CC	-I- DEVELOPMENTAL STAGE: Onset of expression correlates with the				
CC	initiation of axonogenesis in 16-36 hour embryos.				
CC	-I- SIMILARITY: Belongs to the immunoglobulin superfamily.				
CC	LI/neurofascin/NG2CAM family.				
CC	-I- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.				
CC	-I- SIMILARITY: Contains 5 fibronectin type III domains.				

```
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X89204; CAA61490.1; -.
DR PIR; T30581; T30581.
DR HSSP; P20241; 1C8B.
DR ZFIN; ZDB-GENE-980526-512; nad11.1.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FnIII subd.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_C2.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; Ig; 6.
DR PRINTS; PR00014; ENTPEP11.
DR SMART; SM00060; FN3; 5.
DR PROSITE; PS50835; IG_Like; 6.
DR Neurogenesis; Cell adhesion; Developmental protein; Glycoprotein;
DR Transmembrane; Repeat; Immunoglobulin domain.
KW NON TER
FT DOMAIN 1 1054 EXTRACELLULAR (POTENTIAL).
FT TRAMEM 1055 1075 POTENTIAL.
FT DOMAIN 1076 1197 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1 58 IG-LIKE C2-TYPE 1.
FT DOMAIN 165 160 IG-LIKE C2-TYPE 2.
FT DOMAIN 165 263 IG-LIKE C2-TYPE 3.
FT DOMAIN 268 355 IG-LIKE C2-TYPE 4.
FT DOMAIN 360 442 IG-LIKE C2-TYPE 5.
FT DOMAIN 451 541 IG-LIKE C2-TYPE 6.
FT DOMAIN 546 638 FIBRONECTIN TYPE-III 1.
FT DOMAIN 645 739 FIBRONECTIN TYPE-III 2.
FT DOMAIN 744 849 FIBRONECTIN TYPE-III 3.
FT DOMAIN 850 948 FIBRONECTIN TYPE-III 4.
FT DOMAIN 952 1029 FIBRONECTIN TYPE-III 5.
FT DISULFID 92 143 BY SIMILARITY.
FT DISULFID 199 247 BY SIMILARITY.
FT DISULFID 289 339 BY SIMILARITY.
FT DISULFID 383 432 BY SIMILARITY.
FT DISULFID 472 525 BY SIMILARITY.
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 119 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 438 438 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 959 959 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 968 968 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1002 1002 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1027 1027 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1197 AA; 132860 MW; 7CB1505EEFCT828 CRC64;
-----
Query Match 5.1%; Score 117.5; DB 1; Length 1197;
Best Local Similarity 18.0%; Pred. No. 2.7;
Matches 83; Conservative 84; Mismatches 172; Indels 123; Gaps 19;
-----
Oy 9 HLLVLQLALPAAQGNKVVLGKGDYVELTCTASQKKSIQFMKNSNQIKLGNQGSF 68
Db 259 HVTIVTEAA--EYWRSPSEBHLVAPGETVRDLCKADGIPAPITW-SINGVPSGTD--- 312
Oy 69 LTKGSKLNDRAISRSLMDQGNFPLIKLKIETSDTYICEVEDQKEVQLVWGLTAN 128
Db 313 -----VDPKRRV---SSGKLLISNVEFSDTAVYQCAVAKGSIIL-----N 351
```

```
Oy 129 SDTHLQGSILTLT-----LESPPGSSPSVQCRS-----PRGNIGCGK 167
Db 352 THYHVELPQILITPDLRLVQATAGCTVMDLDRFGSPPLPKHMEILDSIPALSNAKISQ 411
Oy 168 T---LSVQLELQDSGTWTCTYLQNKQKVEFKIDV-----VLAFOQASSIVY 212
Db 412 TTNGSLKISNVEEDSNRYTCSVETKKSISADVEVNRKTVGPDPNVLHYIRGSDILH 471
Oy 213 KKEGEQVEFSFPLAFTVEKLTGSGELVWQALERASSKSWITFDLKNKEVSKRYTQDPK 272
Db 472 CK-----YTVDHLKSPTYVMNKDGHKITS--TSNDKHEIEGSKLVLDVQM 517
Oy 273 Q-WCKKLPLHLTLTPQALPOYAGSNTLTALAKTKL-----HQEVNLVWNRATQLO 323
Db 518 EDWG-----IYSCVESTTLDSDTASGYITVQDKPDPQSLKL-----SEKME 559
Oy 324 KNLTCFVWGPV---SPKMLSLKLEKKEKAKVSKREKPVWLVLPKAGMQLCLSDSQVL 379
Db 560 RSVTIS-WMPSVENNSVTEYVIEWEGE-----TPDEQWQKYSRSQDID 605
Oy 380 LESNIXVLPTWSPVHPRASALPAPPTGSLAPPQTSALPD 421
Db 606 SMRSICSYSKTHPOIRAVNSIGTSAPTESSL-SYSTPAAPKD 646
```

```
RESULT 52
C166_CARAV STANDARD; PRT; 555 AA.
ID C166_CARAV
AC Q90304;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE CD166 antigen homolog precursor (Neuroilin) (DM-GRASP homolog).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OX Cyprinidae; Carassius.
ON NCBI_Taxid=7957;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Laessing U.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 30-552 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Retina;
RX MEDLINE=94299040; Pubmed=8026643;
RA Laessing U., Giordano S., Stecher B., Lottspeich F., Stuermer C.A.O.;
RT "Molecular characterization of fish neuroilin: a growth-associated
RT cell surface protein and member of the immunoglobulin superfamily in
RT the fish retinocentral system with similarities to chick protein
RT DM-GRASP/SC-1/BEN-1".
RL Differentiation 56:21-29 (1994).
CC -1- FUNCTION: CELL ADHESION MOLECULE. INVOLVED IN NEURITE EXTENSION BY
CC NEURONS VIA HETEROPHILIC AND HOMOPHILIC INTERACTIONS.
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
CC -1- TISSUE SPECIFICITY: PRESENT ON ALL RETINAL GANGLION CELLS (RGCs)
CC AND THEIR AXONS (IN EMBRYO). ABSENT FROM MATURE AXONS ALONG MOST
CC OF THEIR LENGTH, BUT IS PRESENT ON NEW AND GROWING AXONS DERIVED
CC FROM THE RGCS AT THE RETINAL MARGIN. REMAINS ON ADULT RGCs ONLY AT
CC CELL-CELL CONTACT SITES AND IS CONTINUOUSLY FOUND IN THE RETINAL
CC AXON TERMINAL ARBOR LAYERS OF THE ADULT TECTUM.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like V-type domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
```

```

DR EMBL, L25056; AAC38015.2; -.
DR PIR, I50478; I50478.
DR HSSP, Q13740, I3K3C.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00406; Ig; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
DR Signal; Developmental protein; Cell adhesion; Immunoglobulin domain;
KM Repeat; Glycoprotein; Transmembrane.
FT SIGNAL 1 22
FT CHAIN 23 555
FT DOMAIN 23 555
FT TRANSMEM 500 520
FT DOMAIN 521 555
FT DOMAIN 31 127
FT DOMAIN 131 229
FT DOMAIN 239 323
FT DOMAIN 319 397
FT DOMAIN 406 484
FT DISULFID 38 110
FT DISULFID 154 217
FT DISULFID 263 306
FT CARBOHYD 92 470
FT CARBOHYD 171 171
FT CARBOHYD 350 350
FT CARBOHYD 441 441
FT CARBOHYD 465 465
SQ SEQUENCE 555 AA; 60371 MW; 5A4A8014F00BF68 CRC64;

Query Match 5.0%; Score 116; DB 1; Length 555;
Beet Local Similarity 19.0%; Pred. No. 1.3;
Matches 75; Conservative 69; Mismatches 129; Indels 122; Gaps 20;

QY 29 VLGRKGDVVELTCTASQKKS---IQFMKNSNOIKILNQSGLFKSGK-----LNDK 79
DB 26 VIGLGETIVPCNDGTKKPDGLITTKM--YVDDSPGDLVKAQKQKQKATVSATDG 82
QY 80 ADSRSRLWDQGNFPIIKNLKIEDSDTYICEVEDQKEEVQLVFGILTNSDTHLQGSGL 139
DB 83 YKSRVSI--ANSSILIRAGSLADQVFT-----MVSFTN-----LEESV 123
QY 140 TLTLESPPGSSPVOCSPRGKNIQKGLTSLVQELQDSGTWCTVLQNKRYEFKIDI 199
DB 124 EVKVKKKP-SAPVIK--NNAKELENGKLTQGECEVENA----- 159
QY 200 VVLAQKASIVYKKEGEV-----EFSFPLAFYVEKLTGSGELMWQABRASSKSWITF 254
DB 160 ----NPPADLIWKKNQTLVDGKTIITSTITDKITG-----ISSTSRLOY 204
QY 255 DLKNEKVSRYKVTQPKLQMGKKLPLHLTLPOALQVYAGSGLTLALEAKTGKHL---OE 311
DB 205 TAKKEDVE-SQFTCTAKHYMG--PDQVSEPSFP-----IHYPTK 242
QY 312 VNLVVRATQLOK---NLTCVWGPTSPKMLSLKLENKAKYSKREKPVWL---NP 363
DB 243 VSLQVVSQSPIREGKDVTLKQADGNPP---TSFNFNKKGKAVTVTKDQVYTLTGVTRA 299
QY 364 EAGMWQCLLSDSGVLLSE-----NIKVLPT 389
DB 300 DSGIYKCSLLDND--VMESTQFVTVSFLDVSILPT 332

```

```

DE Neural cell adhesion molecule 1, 140 kDa isoform precursor (N-CAM 140)
GN (NCAM-140).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_Taxid:9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain cortex;
RX MEDLINE=89378239; PubMed=2776887;
RA Lipkin V.M., Khramtsov N.V., Andreeva S.G., Moshayakov M.V.,
RA Petukhova G.V., Rakitina T.V., Reshenko E.A., Ishchenko K.A.,
RA Mitroeva S.F., Chernova M.N., Drianytsyna S.M.;
RT "Calmodulin-independent bovine brain adenylate cyclase. Amino acid
RL sequence and nucleotide sequence of the corresponding cDNA.";
RN FEBS Lett. 254:69-73(1989).
RN [2]
RP SEQUENCE OF 20-36.
RX MEDLINE=86140120; PubMed=3512556;
RA Rougon G., Marshak D.R.;
RT "Structural and immunological characterization of the amino-terminal
RL domain of mammalian neural cell adhesion molecules.";
RN J. Biol. Chem. 261:3396-3401(1986).
RN [3]
RP IDENTIFICATION AS N-CAM.
RX MEDLINE=92111748; PubMed=1765159;
RA Premont R.T.;
RT "A bovine brain cDNA purported to encode calmodulin-insensitive
RL adenylate cyclase has extensive identity with neural cell adhesion
RT molecules (N-CAMs).";
RN FEBS Lett. 295:230-231(1991).
RL -1- FUNCTION: This protein is a cell adhesion molecule involved in
CC neuron-neuron adhesion, neurite fasciculation, outgrowth of
CC neurites, etc.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: Type I isoform=1;
CC Event=Alternative splicing; Named isoform=1;
CC Comment=A number of isoforms are produced;
CC Name=1;
CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
CC -1- CAUTION: Was originally (Ref.1) thought to be a calmodulin-
CC independent adenylate cyclase.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X16451; CA34470.1; -.
DR PIR, A32976; IJBONC.
DR HSSP, P40189; IBCU.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IGc2; 5.
DR PROSITE; PS50835; IG_LIKE; 5.
KM Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KM Immunoglobulin domain; Alternative splicing; Signal.
FT SIGNAL 1 19
FT CHAIN 20 853
FT DOMAIN 20 719
EXTRACELLULAR (POTENTIAL).

```

FT	TRANSMEM	720	737		POTENTIAL.
FT	DOMAIN	728	853		CYTOSOLSMIC (POTENTIAL).
FT	DOMAIN	20	111		IG-LIKE C2-TYPE 1.
FT	DOMAIN	116	205		IG-LIKE C2-TYPE 2.
FT	DOMAIN	212	300		IG-LIKE C2-TYPE 3.
FT	DOMAIN	307	412		IG-LIKE C2-TYPE 4.
FT	DOMAIN	415	500		IG-LIKE C2-TYPE 5.
FT	DOMAIN	527	604		FIBRONECTIN TYPE-III 1.
FT	DOMAIN	633	700		FIBRONECTIN TYPE-III 2.
FT	DOMAIN	152	156		HEPARIN-BINDING (POTENTIAL).
FT	DOMAIN	161	165		HEPARIN-BINDING (POTENTIAL).
FT	DISULFID	41	96		BY SIMILARITY.
FT	DISULFID	139	189		BY SIMILARITY.
FT	DISULFID	235	286		BY SIMILARITY.
FT	DISULFID	328	394		BY SIMILARITY.
FT	DISULFID	435	468		BY SIMILARITY.
FT	CARBOHYD	222	222		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	314	314		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	346	346		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	432	432		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	458	458		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	487	487		N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	853 AA:	93893 MW;	E12FD49231A7A368 CR664;	
<hr/>					
Query Match 5.0%; Score 116; DB 1; Length 853;					
Best Local Similarity 20.5%; Pred. No. 2.2;					
Matches 87; Conservative 52; Mismatches 159; Indels 126; Gaps 18					
OY	32	KKGDVVELCTASQSKSIQFHKNQNIKIIGNQSFLTKPSKLNDRADSRSLMDGN	91		
DB	130	RREGDAIVVCDDVSSLPTTIWKHGRDYTLKKDVRFIV-----	LTN	171	
OY	92	PPLIKNLKIEDSDTYICE-----VEDQKEVOLLV-----	FGLTANSDFHL	133	
DB	172	NYLQIRGIKKTDGEGYRCBGRILANGELINFKDQIYVANVPRTQAQSIVANATAN-----		226	
OY	134	LOGSGLITLESPPSSPSVOCRSRGKRIQC-----GKLTSVSLELDQSGTW	182		
DB	227	LGOSEVTLLVCNAEGFPPEPVSW-TYDGEOINEDEEKKYLFSDDSSELTIRKYDKNDDEAY	284		
OY	183	TCTVLQNOKVEFKIDIVLAFQKASSIYYKKEGEVERFSPLATFEVKLTSGE----	237		
DB	285	VCIANKKAGEQDASHLKVFAPK---ITTVNQIMALEEBGVTLTCE---ASGPDIISI	338		
OY	238	LWMAERASSSK--SWTFEDIKANESVRYTODPKIQMGKRLPL-HITLPALPYAG	293		
DB	339	TWRSTSTRNISSEKASWTRE-----KQETLDGHMVVRSHAVSSYLTKSI--CYTD	388		
OY	294	SGNTLLALEAKTGKHQENLVLMVMAATQ-----QXNLTCFYWGTPSPKMLMS	341		
DB	389	AGEIVCTASNTIGQSQSMYLEVOYAPKIQGVANVYTMEGNOVNITCEVFAYS-----	442		
OY	342	LKLENKAQSVSKREKRVWLNEADGMOCCLSDSQVILLNESNIKVLPTMSPVHPRASAL	401		
DB	443	-----ATIS-----WFRD-----COLLPSYSNYISKIYNT--PSASYL	473		
OY	402	PAPP	405		
DB	474	EVTIP	477		
<hr/>					
RESULT 54					
FPRP RAT STANDARD; PRT; 879 AA.					
ID	FPRP RAT				
AC	Q62786;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Prostaglandin F2 receptor negative regulator precursor (Prostaglandin				
DE	F2-alpha receptor regulatory protein) (Prostaglandin F2-alpha receptor				
DE	associated protein).				
EN	PTGFRN OR FPRP.				
OS	Rattus norvegicus (Rat).				

OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OK	NCBI_TaxId=10116;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Sprague-Dawley; TISSUE=Ovary;
RX	MEDLINE=97109715; PubMed=8951995;
RA	O'licky D.J., Nordeen S.K.:
RT	"Cloning, sequencing and proposed structure for a prostaglandin F2 alpha receptor regulatory protein.", Rt prostaglandins leukot. Essent. Fatty Acids 55:261-268(1996). [2]
RL	
RN	CHARACTERIZATION.
RC	STRAIN=Sprague-Dawley; TISSUE=Ovary;
RX	MEDLINE=96397016; PubMed=8804121;
RA	O'licky D.J.:
RT	"Negative regulatory activity of a prostaglandin F2 alpha receptor associated protein (FPPR)."; Rt prostaglandins leukot. Essent. Fatty Acids 54:247-259(1996).
RL	
CC	-1 FUNCTION: Inhibits the binding of prostaglandin F2-alpha (PGF2- alpha) to its specific FP receptor, by decreasing the receptor number rather than the affinity constant. Functional coupling with the prostaglandin F2-alpha receptor seems to occur. CC -1 SUBCELLULAR LOCATION: Type I membrane protein. Endoplasmic reticulum and trans-Golgi network. CC -1 SPECIFICITY: REPRODUCTIVE TISSUES, LUNG AND HEART. CC -1 SIMILARITY: Contains 6 immunoglobulin-like C2-type domains. CC ----- CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-sib.ch). CC or send an email to license@isb-sib.ch . CC ----- DR EMBL; U26595; AAC18426.1; - DR GO; GO:0005515; F:protein binding; GPI. DR GO; GO:0017148; P:negative regulation of protein biosynthesis; NAS. DR InterPro; IPRO07110; IG-like. DR InterPro; IPRO03596; IG_v. DR Pfam; PF00047; Ig_6. DR SMART; SM00406; IGV_1. DR PROSITE; PS00835; IG-LIKE; 5. KW Signal; Immunoglobulin domain; Transmembrane; Glycoprotein; Repeat; KW Polymorphism. FT FT SIGNAL 1 21 POTENTIAL. FT CHAIN 22 879 PROTAGLANDIN F2 RECEPTOR NEGATIVE REGULATOR. FT FT DOMAIN 22 832 EXTRACELLULAR (POTENTIAL). FT TRANSMEM 833 853 POTENTIAL. FT DOMAIN 854 879 CYTOPLASMIC (POTENTIAL). FT FT DOMAIN 22 129 IG-LIKE C2-TYPE 1. FT DOMAIN 149 268 IG-LIKE C2-TYPE 2. FT DOMAIN 276 389 IG-LIKE C2-TYPE 3. FT DOMAIN 406 527 IG-LIKE C2-TYPE 4. FT DOMAIN 544 662 IG-LIKE C2-TYPE 5. FT DOMAIN 688 813 IG-LIKE C2-TYPE 6. FT DOMAIN 424 427 ENDOLASMIC RETICULAR RETENTION SIGNAL. FT SITE 89 91 CELL ATTACHMENT SITE (POTENTIAL). FT SITE 703 705 CELI ATTACHMENT SITE (POTENTIAL). FT DISULFD 43 119 BY SIMILARITY. FT DISULFD 169 247 BY SIMILARITY. FT DISULFD 299 373 BY SIMILARITY. FT DISULFD 429 515 BY SIMILARITY. FT DISULFD 571 655 BY SIMILARITY. FT DISULFD 711 793 BY SIMILARITY. FT CARBOHYD 44 44 N-LINKED (GLCNAC . .) (POTENTIAL). FT CARBOHYD 300 300 N-LINKED (GLCNAC . .) (POTENTIAL). FT CARBOHYD 383 383 N-LINKED (GLCNAC . .) (POTENTIAL). FT CARBOHYD 413 413 N-LINKED (GLCNAC . .) (POTENTIAL). FT CARBOHYD 525 525 N-LINKED (GLCNAC . .) (POTENTIAL). FT CARBOHYD 600 600 N-LINKED (GLCNAC . .) (POTENTIAL).

```
FT CARBOHYD 618 618 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 691 691 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 47 47 D -> G.
FT VARIANT 136 136 M -> V.
FT VARIANT 782 782 S -> G.
FT VARIANT 844 844 G -> R.
SQ SEQUENCE 879 AA; 98730 MW; DE7012D3B346C0F7 CRC64;

Query Match
Beet Local Similarity 18.6%; Pred. No. 2.5;
Matches 87; Conservative 78; Mismatches 175; Indels 129; Gaps 18;

14 LQALPAPATQGNKVLGGKDTVELTCTASQ-----KKSQPFMKNSNQKILG----- 63
14 LQALV-----PRVSVTEGKDLDSCTITDRDDVDVPEVTWTFKPTPTSLASHML 331
64 ---NGSFLTKGPSKLNDRADSRSLMDGNFPLIINKLIEDSDTYICEV-----ED 113
332 ARLDRLSLVSSPHVALSHVDTR-----SYHLVRDVSKEKNGVYLCVLAMPGNHR 384
114 QKEEVQLVFLGLTANSQDTHLLQGSILTLTLESPPGSS---PSVQCRSPRGKNIQGGKTL 170
385 SMHVAEMASAPSGVSVTLBEPEGYVLNASKVPFSPDPTELQCRVIDTKKVDAGVRLT 444
171 V-----SOLEQDSGWTCTVLQONKKVEFKIDIVLAFQKA--SIIYK 213
445 VSMYRMRNRNDVVASSELLAMVDGWTLYRGERSK-----QNRQDSGPFIS 491
214 KEGQVEFSPLAFTVEKLTGSGELMWAQERASSSKSWI--TPDLNKEVSVAKVTQDPKL 272
492 KENTDT-FSFRIGRTTEEDRGSGYCVASMTQRNNSWKSADVPSKPNIMFASGDSVL 550
273 QMGKPLPLHLTPALPOYASGSLTLLEATGKTLHGVNLVVRATQLOKNLTCEWIG 332
551 VYKAPQ-----KPFPAAGNTFEMTCKVSSKNIKSPRYVLTITAEKPVGDL-----S 598
333 PTPSKLMLSL-----KLEN-----KEAKYSKREKPPVWNLNPEAGMOCILS 373
599 PNEYTIISLDSDSVKLEKMTDASRVDCVLEKQVDEFFRYRMVQTOVSDGLYRCMT 658
374 DSGQVLLBSNIVKLPVTPVPRASALPAPPTGSAIPDPPTASALPP 422
659 -----AMS-----PIGSL-WREKATSLSN 678

RESULT 55
TIE1_BOVIN
ID TIE1_BOVIN STANDARD; PRT; 1136 AA.
AC 006805;
DT 01-FEB-1995 (rel. 31, Created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 28-FEB-2003 (rel. 41, Last annotation update)
DE Tyrosine-protein kinase receptor TIE-1 precursor (EC 2.7.1.112).
GN TIE OR TIE1 OR TIE-1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Kuminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endothelial cells;
RX MEDLINE=9402374; PubMed=8415706;
RA Sato T.N., Qin Y., Kozak C.A., Andrus K.L.;
RT "Tie-1 and tie-2 define another class of putative receptor tyrosine
RT kinase genes expressed in early embryonic vascular system.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:9355-9358(1993).
CC -1- FUNCTION: Probable protein tyrosine-kinase transmembrane receptor.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Specifically expressed in developing vascular
CC endothelial cells.
```

```
CC -1- SIMILARITY: Belongs to the Tyr family of protein kinases. Tie
CC subfamily.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 3 fibronectin type III domains.
CC -1- SIMILARITY: Contains 3 EGF-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X71423, CA50554.1; -.
CC PIR, S57845; S57845.
CC DR HSSP; P11362; IFGK.
CC DR InterPro; IPR006209; EGF like.
CC DR InterPro; IPR008957; FN_III-like.
CC DR InterPro; IPR003961; FN_III.
CC DR InterPro; IPR006210; IEGF.
CC DR InterPro; IPR007110; IG-like.
CC DR InterPro; IPR003599; IG.
CC DR InterPro; IPR002049; Laminin EGF.
CC DR InterPro; IPR000719; Prot. kinase.
CC DR InterPro; IPR001245; Tyr_kinase.
CC DR InterPro; IPR008266; Tyr_kinase_AS.
CC DR Pfam; PF00008; EGF; 2.
CC DR Pfam; PF00041; Fn3; 3.
CC DR Pfam; PF00047; Ig; 2.
CC DR Pfam; PF00069; pkinase. 1.
CC DR PRINTS; PR00109; TYRKINSE.
CC DR PRINTS; PR00109; TYRKINSE.
CC DR ProDom; PD000001; Prot_kinase; 1.
CC DR SMART; SM00181; EGF; 2.
CC DR SMART; SM00060; FN3; 3.
CC DR SMART; SM00409; IG; 1.
CC DR SMART; SM00219; TYKIC; 1.
CC DR PROSITE; PS00022; EGF_1; 3.
CC DR PROSITE; PS01186; EGF_2; 3.
CC DR PROSITE; PS50026; EGF_3; 2.
CC DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC DR Receptor, Tyrosine-protein kinase; Transferase; Signal; ATP-binding;
CC Repeat; EGF-like domain; Transmembrane; Immunoglobulin domain;
CC Glycoprotein; Phosphorylation; Multigene family.
CC KW SIGNAL.
CC FT CHAIN 1 24 1136 TYROSINE-PROTEIN KINASE RECEPTOR TIE-1.
CC FT DOMAIN 24 757 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 758 782 POTENTIAL.
CC FT DOMAIN 783 1136 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 43 106 IG-LIKE C2-TYPE 1.
CC FT DOMAIN 212 254 IG-LIKE 1.
CC FT DOMAIN 256 301 EGF-LIKE 2.
CC FT DOMAIN 303 343 EGF-LIKE 3.
CC FT DOMAIN 370 424 IG-LIKE C2-TYPE 2.
CC FT DOMAIN 444 538 FIBRONECTIN TYPE-III 1.
CC FT DOMAIN 541 637 FIBRONECTIN TYPE-III 2.
CC FT DOMAIN 641 742 FIBRONECTIN TYPE-III 3.
CC FT DOMAIN 837 1116 PROTEIN KINASE.
CC FT NP_BIND 843 851 ATP (BY SIMILARITY).
CC FT BINDING 868 868 ATP (BY SIMILARITY).
CC FT ACT_SITE 977 977 BY SIMILARITY.
CC FT MOD_RES 1005 1005 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 707 707 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 1136 AA; 124953 MW; EEF805804A041BB12 CRC64;

Query Match
5.0%; Score 115.5; DB 1; Length 1136;
```

Best Local Similarity 21.4%; Pred. No. 3.5; Matches 86; Conservative 47; Mismatches 146; Indels 123; Gaps 18

```
OY 109 CEVEDQKEEVQLLVGLTANSDDLHQQOSLTLTLESPPGSP-----SVQCRSPGKNI 163
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 342 CEKSDRIPQILDVMSSELEFNLDT-----MPRINCAAGNPFVPGSMELRKDGTVL 393
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 164 QGKGLT-----SVSOLEDSGWTCTVLQN--QKKVEKIDIVLAFQKASSTV 211
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 394 LSTAIIVEDPRTTAEFEVPRIALGSGMECKVSTSGQDSRRFRINVKVPPVLTARL 453
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 212 YKKEGEQVEFSFPLAFTVEKLTGSGELMWQAEKASSKSWITFDLKNKEVSKVATODPK 271
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 454 LAKSROLVSPVSPVSGSDGPIASVRLHYRRP--DSTYAMST-----IVDP- 498
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 272 LQMGKPLPLHLTLPGALPQVAGSGNLTLA-LEAKTGKLHQEVLVWMPATQLQKNLTCEV 330
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 499 -----SENVTLMINLRPKTG---YSVRVQLSRPGEGEG-----A 529
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 331 WGPPTS-----PKLMLSLKLEKKAKEKSKREPVVNLVPEAG-----MQCLISD 374
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 530 WGPPLTLMTTDCPEPLKPKWLEGMHVEGPDRLRVSWSLPPVPGLVGDGFLRLMD---GA 586
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 375 SGQVLTLESNIK-----VLPTWST-----PVHPRASALPAPPTGSAL 410
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 587 RQGERRE-NVSSPQARKTALLGTLTGTYQLDVRVLYHCTLGLGSPAPPAARVL-LPESGPPA 644
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 411 PDPQTASALPDP-----PAASALPALAVISFLGLGLG 444
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 645 PRLHLAQLSDSEIQLMWQRPAAAGPISKYIVVEQVAGSGG 686
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

RESULT 56

NRG_DROME STANDARD; PRT; 1302 AA.

AC P20241; O61541; O61542; Q24414; Q24415; Q95064; Q9V3X0;

DT 01-FEB-1991 (Rel. 17, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Neuroglian precursor.

GN NRG OR CG1634.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; OC Ephydroidea; Drosophilidae; Drosophila.

CC NCBI_TaxID=7227;

OX [1]

RP SEQUENCE FROM N.A. (ISOFORM SHORT), AND SEQUENCE OF 24-41 AND 737-751.

RN MEDLINE=90030418; PubMed=2805067;

RX Biedler A.J., Snow P.M., Hortsch M., Patel N.H., Jacobs J.R.,

RA Tiedjema Z.R., Schilling J., Goodman C.S.;

RT "Drosophila neuroglian: a member of the immunoglobulin superfamily

RT with extensive homology to the vertebrate neural adhesion molecule

RT L1.";

RL Cell 59:447-460(1989).

RN [2]

RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RX MEDLINE=98332718; PubMed=9666073;

RA Zhao G., Hortsch M.;

RT "The analysis of genomic structures in the L1 family of cell adhesion

RT proteins provides no evidence for exon shuffling events after the

RT separation of arthropod and chordate lineages.";

RL Gene 215:47-55(1998).

RN [3]

RP REVISIONS.

RA Hortsch M.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A. (ISOFORM SHORT).

RC STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.
RA Suton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
RA Abin K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mkilos G.L.G.,
RA Avril J.F., Agbayanti A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Baau A.U., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benson P.V., Bertman B.P., Bhendari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Butts K.C., Buam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke J., Davernport L.B., Davies P.,
RA de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabriellan A.E., Gang N.S., Gelbart W.M., Glaesser K.,
RA Gloder A., Gong F., Gorrell U.H., Gu Z., Guan P., Harris M.,
RA Harris N.D., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
RA Jalali M., Kalush F., Kapzen G.H., Ke Z., Kennison J.D., Kerchum K.A.,
RA Kammeil B.E., Kodira C.D., Kraft C., Kravitz S., Kulj D., Lai Z.,
RA Laako P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekullov G., Mishina N.V., Mobarry C., Morris J., Mosheiff A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzy D.M., Nelson D.L.,
RA Neilson D.R., Nielson K.A., Nixon K., Nysskern D.R., Paclel J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spitzer E., Spradling A.C., Stepieton M., Strong R., Sun E.,
RA Sylvester R., Teetor C., Turner R., Ventner E., Wang A.H., Wang X.,
RA Wang Z.-Y., Masaramba D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zavatti J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,
RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2195-2195(2000).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC STRAIN=Berkely; TISSUE=Head;
RA Stepleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champé M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarín H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Paclel J., Paragas V., Park S., Phouanavongs S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/genbank/DBJ databases.
RN [6]
RP SEQUENCE OF 1182-1302 FROM N.A., FUNCTION, ALTERNATIVE SPLICING, AND
RP TISSUE SPECIFICITY.
RC TISSUE=Embryo;
RA Hotteich M., Bieber A.J., Patel N.H., Goodman C.S.;
RX PubMed=90262720; PubMed=1693086;
RA Hotteich M., Bieber A.J., Patel N.H., Goodman C.S.;
RT "Differential splicing generates a nervous system-specific form of
RL *Drosophila neuroligin*.";
RL Neuron 4:697-709(1990).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 610-814.
RX PubMed=94231741; PubMed=7512815;
RA Huber A.H., Wang Y.-M.E., Bieber A.J., Bjorkman P.J.;
RL "Crystal structure of tandem type III fibronectin domains from
RT *Drosophila neuroligin* at 2.0 Å.";
RL Neuron 13:717-731(1994).
CC -I- FUNCTION: The long isoform may play a role in neural and glial
cell adhesion in the developing embryo. The short isoform may be
more general cell adhesion molecule involved in other tissues and
imaginal disk morphogenesis. Vital for embryonic development.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P20241-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P20241-2, Sequence=VSP_002601, VSP_002602,
CC -I- TISSUE SPECIFICITY: LONG ISOFORM IS RESTRICTED TO SURFACE OF

CC NEURONS AND GLIA IN THE DEVELOPING NERVOUS SYSTEM AND THE SHORT
CC ISOFORM TO OTHER NONNEURONAL TISSUES.
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 5 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M28231; AAA28728.2; -
DR EMBL: AF050085; AAC28613.2; JOINED.
DR EMBL: AF050084; AAC28614.2; -
DR EMBL: AF050085; AAC28614.2; -
DR EMBL: AE003444; AAF46387.1; -
DR EMBL: AY058284; AAL13513.1; -
DR EMBL: X76243; CA533822.1; -
DR EMBL: X76244; CA533823.1; -
DR PIR: A32579; A32579.
DR PDB: 1CFB; 30-NOV-94.
DR FLYbase; FBgn002968; Nrg.
DR GO: GO:0005886; C:plasma membrane; IMP.
DR GO: GO:0005194; P:cell adhesion molecule activity; IMP.
DR GO: GO:0007560; P:imaging disc morphogenesis; IMP.
DR GO: GO:0007158; P:neutrophil cell adhesion; IMP.
DR InterPro: IPR008957; FN III-like.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003598; IG_c2.
DR Pfam: PF00041; fn3; 5.
DR Pfam: PF00047; Ig; 6.
DR SMART; SMO0060; FN3; 5.
DR SMART; SMO0408; IGC2; 4.
DR PROSITE; PS00835; IG_Like; 6.
DR Cell adhesion; Glycoprotein; Transmembrane; Repeat; 3D-structure;
KM Immunoglobulin domain; Signal; Developmental protein;
KM Alternative splicing.
KM SIGNAL 1 23
FT CHAIN 24 1302
FT DOMAIN 24 1138
FT TRASMEM 1139 1154
FT DOMAIN 1155 1302
FT DOMAIN 29 133
FT DOMAIN 134 225
FT DOMAIN 245 330
FT DOMAIN 339 426
FT DOMAIN 432 524
FT DOMAIN 521 610
FT DOMAIN 629 690
FT DOMAIN 729 792
FT DOMAIN 832 896
FT DOMAIN 932 997
FT DOMAIN 1024 1098
FT DISULFID 59 111
FT DISULFID 625 706
FT CARBOHYD 182 182
FT CARBOHYD 188 198
FT CARBOHYD 411 411
FT CARBOHYD 448 448
FT CARBOHYD 652 652
FT CARBOHYD 663 663
FT CARBOHYD 821 821
FT CARBOHYD 1125 1125
FT VARSPLIC 1224 1239
FT VARSPLIC 1240 1302
FT CONFLICT 85 86
NR -> KP (in REF. 2).

FT CONFLICT 1282 1282 MISSING (IN REF. 6).
FT STRAND 619 625
FT STRAND 629 635
FT TURN 640 641
FT STRAND 646 653
FT TURN 657 658
FT STRAND 661 668
FT TURN 669 670
FT STRAND 673 677
FT STRAND 682 692
FT TURN 693 694
FT STRAND 695 696
FT STRAND 706 708
FT STRAND 721 723
FT TURN 727 728
Query Match 5.0%; Score 115.5; DB 1; Length 1302;
Best Local Similarity 19.7%; Pred. No. 4.2;
Matches 67; Conservative 61; Mismatches 129; Indels 83; Gaps 16;
QY 25 GNKVVLGKKDPTVELTCTASQKK-SIOFHKNSQIKILNQGS-FLTKGPSKLNDR- 80
Db GNKVVLL---DVKQGVASQNKHPVQYVSRQSLRQKRMELFCIYGTPLPQTVM 280
QY 81 --DSRRSLMD---QGNF--PLIINKLIEDSDTYCEVEDQKEVOLLVFGLTANSPT 132
Db SKDGRIRMSRIRDTGSHGKSLVRQTFDDAGTYTCDVSGVNAQSFIIILNVNSVPY 340
QY 133 LLOGSLTLTLSPGSSPSVQCRS---PRGKNIQCGK-----TLNV 171
Db LTKKEPELNTAED---EEVFECPRAAGVPEPKISWIHGKLEIGSTPNPRRTVDTIRI 397
QY 172 SQLELDGSGTWTCTYQLQ---QKVEFKIDIVLAFAKASISYKKGEOVESFPPLAF 227
Db INLVGDDTGNVCNATNSIGVYKDVLYLVQAEPPTISEAPAAVSTVDGRNV----- 449
QY 228 TVE-KLTGSGELMNOAEPAASSKSWIT--FDLK-NKEVSRYRTQDPKLGKKLPLHL 282
Db TIKCRVNSPPLVYKMLRAS---NMLTGRYVNOAGDLEIQDVT----- 491
QY 283 TLPPQALPYAGSGNLTLLAEKTKGLHDEVNLVVRATOL 322
Db 492 -----FSDAGKYTCYAQNKFGELQADGSLVYKHEIRI 523
RESULT 57
FCEA HUMAN STANDARD; PRT; 257 AA.
ID FCEA HUMAN
AC P12319;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE High affinity immunoglobulin epsilon receptor alpha-subunit precursor
DE (FceRI) (IGE Fc receptor, alpha-subunit) (Fc-epsilon RI-alpha).
GN FCERIA OR FCEA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN 11
RP SEQUENCE FROM N.A..
RX MEDLINE=86233953; Pubmed=2967464;
RA Kochan J., Petline L.F., Hakim J., Kishi K., Kinet J.-P.;
RT "Isolation of the gene coding for the alpha subunit of the human high
RL affinity IGE receptor."
RL Nucleic Acids Res. 16:3584-3584(1988).
RN 12
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Heart cells;
RX MEDLINE=88158102; Pubmed=2964640;
RA Shimizu A., Tepfer I., Bentley P.N., Berenstein E.H., Sireganian R.P.,
RA Leder P.;
RT "Human and rat mast cell high-affinity immunoglobulin E receptors:


```

RN [2]
RP REVISIONS.
RA Kochan J.;
RL Submitted (MAR-1988) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Maest cells;
RX MEDLINE=88158102; PubMed=2964640;
RA Shimizu A., Tepler I., Bentley P.N., Berenstein E.H., Stragmanian R.P.,
  Leder P.;
RT "Human and rat mast cell high-affinity immunoglobulin E receptors:
  characterization of putative alpha-chain gene products.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:1907-1911(1988).
RN [4]
RP SEQUENCE OF 21-245 FROM N.A.
RX MEDLINE=88289772; PubMed=2969594;
RA Liu F.-T., Albrandt K., Robertson M.W.;
RT "CDNA heterogeneity suggests structural variants related to the high-
  affinity IgE receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:5639-5643(1988).
CC -1- FUNCTION: BINDS TO THE FC REGION OF IMMUNOGLOBULIN EPSILON. HIGH
  AFFINITY RECEPTOR. RESPONSIBLE FOR INITIATING THE ALLERGIC
  RESPONSE. BINDING OF ALLERGEN TO RECEPTOR-BOUND IgE LEADS TO CELL
  ACTIVATION AND THE RELEASE OF MEDIATORS (SUCH AS HISTAMINE)
  RESPONSIBLE FOR THE MANIFESTATIONS OF ALLERGY. THE SAME RECEPTOR
  ALSO INDUCES THE SECRETION OF IMPORTANT LYMPHOKINES.
CC -1- SUBUNIT: TETRAMER OF AN ALPHA CHAIN, A BETA CHAIN, AND TWO
  DISULFIDE LINKED GAMMA CHAINS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M17153; AAA42045.1; -
DR EMBL, J03606; AAA41582.1; -
DR EMBL, M21622; AAA41146.1; -
DR PIR, C31327; A30154.
DR HSSP; P12319; IALS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003599; Ig.
DR Pfam; PF00047; Ig_2.
DR SMART; SM00409; Ig; 2.
DR PROSITE; PS00835; IG_LIKE; 1.
KW IgE-binding protein; Receptor; Transmembrane; Glycoprotein; Signal;
  Immunoglobulin domain; Repeat.
FT CHAIN 1 23
FT SIGNAL 24 245
FT DOMAIN 24 245 HIGH AFFINITY IMMUNOGLOBULIN EPSILON
  RECEPTOR ALPHA-SUBUNIT.
FT DOMAIN 205 223 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 224 245 POTENTIAL.
FT DOMAIN 28 103 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 113 181 IG-LIKE 1.
FT DISULFID 49 91 BY SIMILARITY.
FT DISULFID 130 174 BY SIMILARITY.
FT CARBOHYD 52 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 245 AA; 27793 MW; A0867DD363B72197 CR664;
Query Match 4.9%; Score 113.5; DB 1; Length 245;
Best Local Similarity 23.5%; Pred. No. 0.66;
Matches 63; Conservative 41; Mismatches 85; Indels 79; Gaps 14;

```

```

QY 11 LVLQLALLPAATGQNKVVLGK-----KQDVELTC---TASQKSIQPFHKNKSNQIKI 61
DB 13 LVLISLGMLTATKQKSVSLDPPIRITITGKATLTCGNKNSQMNSTK--WINDSI-- 68
QY 62 LGNGSPLTKGPKSLNDRSRSLMDQGNPFLIKIKLIEDSDTYICEVED--QKEEVQ 119
DB 69 -----SNVKSASH-----YIVSATIIDSGKVIQKGFYKSKFVY 103
QY 120 LLYGLTANSDTHLLQGOSLTLTIESPPGSPSYQCSPPG-----KNIQ 164
DB 104 LNMV-----QEWLLQSSADVLDN---GSFDIRCSRWKWKVHYKYKDIAFKSY 154
QY 165 GKTLSVQSLELODSGTWTCTVLONQKVEFKIDIVLAQKASIVYKKEGEVFEFP 224
DB 155 DSNINISIKATPNDSGTHCTGYLN--KVECKSDKFSIAYVKDYTIETR---WLQLIFP 208
QY 225 ----LAFTVEKLTGSGELMWAERASS 248
DB 209 SLAVILFAVD--TG---LWFSTHQFES 231

RESULT 59
LAMP CHICK STANDARD; PRT; 338 AA.
AC 096919;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Limbic system-associated membrane protein precursor (E19S) (CHLAMP,
  G19-isoform).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
  OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97157768; PubMed=9004047;
RA Wilson D.J.A., Kim D.S., Clarke G.A., Marshall-Clarke S., Moss D.J.;
RT "A family of glycoproteins (GP55), which inhibit neurite outgrowth,
  are members of the Ig superfamily and are related to OBGM,
  RT neurotrophin, LAMP and CEP-1.";
RL J. Cell Sci. 109:3129-3138(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97358596; PubMed=9215692;
RA Brummendorf T., Spaltmann F., Treubert U.;
RT "Cloning and characterization of a neural cell recognition molecule
  on axons of the retinorectal system and spinal cord.";
RL Eur. J. Neurosci. 9:1105-1116(1997).
CC -1- FUNCTION: Mediates selective neuronal growth and axon targeting.
  Probably serves as a recognition molecule for the formation of
  limbic connections (By similarity)
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
  similarity).
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. IGLON
  family.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL, Y08171; CAA69357.1; -
DR EMBL, Z94720; CAB08115.1; -
DR InterPro; IPR007110; Ig-like.

```

DR InterPro; IPR003598; Ig_c2.
 DR Pfam; PF00047; Ig_3.
 DR SMART; SM00408; IGC2; 2.
 DR PROSITE; PS50835; IG_LIKE; 3.
 KW Immunoglobulin domain; Cell adhesion; Glycoprotein; Lipoprotein;
 KW Membrane; GPI-anchor; Repeat; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 315
 FT PROPEP 316 338
 FT DOMAIN 29 122
 FT DOMAIN 132 214
 FT DOMAIN 219 306
 FT DISULFID 53 111
 FT DISULFID 153 197
 FT DISULFID 239 290
 FT LIPID 315 315
 FT CARBOHYD 40 40
 FT CARBOHYD 66 66
 FT CARBOHYD 136 136
 FT CARBOHYD 148 148
 FT CARBOHYD 279 279
 FT CARBOHYD 287 287
 FT CARBOHYD 300 300
 FT CARBOHYD 315 315
 SQ SEQUENCE 338 AA; 37394 MW; 8FA4A60AD9842684 CRC64;

Query Match 4.9%; Score 113.5; DB 1; Length 338;
 Best Local Similarity 21.0%; Pred. No. 1;
 Matches 83; Conservative 64; Mismatches 144; Indels 105; Gaps 20;

OY 10 LLLVQLALLPAA-----TOGKVVLGKKGDTVELTCTASOKKSIQFMKKNNOIKI 61
 DB 14 LVLRLLCLLFTGCPVRSVDFTRGDITVAGOGTALRCVEDESS-KVAMLRSGIIF 72
 OY 62 LGNGGSLTQSPKSLNDRAD-SRSLMDQGNPFLIKKLIKEDSDTYICEVEDQ---KE 116
 DB 73 AG-----EDKWSIDPRVELEKSPLE---YSLRIQVDVYDEGSYCSVQTHHPKTS 122
 OY 117 EVOLLVFG-----LTANSTHLLQGSLLTLESPPGSSPSVQCR--SRGKNIQGGT-L 169
 DB 123 QVYLIVQVPKISNISDITVNEGSNVTLVCMANGRPVITWRHLPTGGEFGESEBYL 182
 OY 170 SVSOLELDGSGTWCTVQNGKQKVEFKIDIVLAFQKASSIYVKKGEQVQ--PSPLAF 227
 DB 183 EILGTRGSGKYE-----KANEVASADVQVAVYVNYPTI 221
 OY 228 TVEKLTGSGELMWAERASSKSWITFDLKKKEVSV-----KRVTDPKLQMGKCLPLH 281
 DB 222 TESK-----SNEAATGRQA-----LLRCEASAVPTPDPEWRYDDTRINSANGLEIK 267
 OY 282 LTLPGALPOYAGSGULTLALAKTKGLHGVNLVVMRTQIQKULTCVMWPTSPKML- 340
 DB 268 ST-----SQSILMVAANYTEHYG-NYTCVANKL-----GVNASLYLY 306
 OY 341 ---SLKENKEAKYSKREKPVNLNPEAGMOCILS 373
 DB 307 RPTGRVNGSVSILA---VPLML---AASILCLIS 336

RESULT 60
 CAML_FUGRU
 ID CAML_FUGRU STANDARD; PRT; 1277 AA.
 AC Q98902;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Neutral cell adhesion molecule L1 precursor (N-CAM L1) (L1-CAM).
 GN L1CAM.
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Takifugu.
 OX NCBI_Taxid=31033;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Brain, and Muscle;
 RX MEDLINE=9814798; PubMed=9479034;
 RA Coutelle O., Nyakatura G., Taudien S., Elgar G., Brenner S.,
 RA Platzer M., Drescher B., Jonet M., Kenwright S., Rosenthal A.;
 RT "The neutral cell adhesion molecule L1: genomic organisation and
 RT differential splicing is conserved between man and the pufferfish
 RT Fugu";
 RL Gene 208:7-15 (1998).
 CC -1- FUNCTION: Cell adhesion molecule with an important role in the
 CC development of the nervous system. Involved in neuron-neuron
 CC adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds
 CC to axonin on neurons.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=Brain;
 CC IsoId=Q98902-1; Sequence=Displayed;
 CC Name=2; Synonyms=Muscle;
 CC IsoId=Q98902-2; Sequence=VSP_050474, VSP_050475;
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.
 CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 5 fibronectin type III domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z71926; CA96469.1; -;
 DR EMBL; AF026198; AAC15580.1; -;
 DR PIR; T30532; T30532.
 DR HSP; P20241; 1CFB.
 DR InterPro; IPR008957; FN_III-like.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003598; Ig_c2.
 DR Pfam; PF00041; fn3; 5.
 DR Pfam; PF00047; Ig_5.
 DR SMART; SM00060; FN3; 5.
 DR SMART; SM00408; IGC2; 3.
 DR PROSITE; PS50835; IG_LIKE; 6.
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat;
 KW Immunoglobulin domain; Signal; Alternative splicing.
 FT SIGNAL 1 34
 FT CHAIN 35 1277
 FT DOMAIN 30 1135
 FT TRANSMEM 1136 1156
 FT DOMAIN 1157 1277
 FT DOMAIN 1 140
 FT DOMAIN 150 241
 FT DOMAIN 256 344
 FT DOMAIN 349 437
 FT DOMAIN 443 528
 FT DOMAIN 532 623
 FT DOMAIN 628 726
 FT DOMAIN 727 825
 FT DOMAIN 826 932
 FT DOMAIN 933 1028
 FT DOMAIN 1029 1122
 FT DISULFID 129 224
 FT DISULFID 173 224
 FT DISULFID 280 328
 FT DISULFID 370 421
 FT DISULFID 465 514
 FT DISULFID 465 514

```

FT DISULFID 554 607 BY SIMILARITY.
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 794 794 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 839 839 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1035 1035 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1046 1046 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1068 1068 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1083 1083 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1108 1108 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT VARSPLIC 42 47 YVSDL -> F (in isoform 2).
FT VARSPLIC 1190 1193 Missing (in isoform 2).
SQ SEQUENCE 1277 AA; 141954 MW; 2848B49BA9A42C27 CRC64;
Query Match 4.9%; Score 113.5; DB 1; Length 1277;
Best Local Similarity 21.7%; Pred. No. 5.6; Mismatches 152; Indels 81; Gaps 16;
Matches 80; Conservative 56;
QY 34 GDTVELTCTASQKSIQFHKNSNOIKILGNQSFLLTKGPKSLNDRADSRSLMDQGNFP 93
DB 363 GETVLKDCQADGIPSPITW-----TVNGVPLSATSLPRRSLTESGS-- 405
QY 94 LIKKLKIEDSTYICEVEDQEE-----VOLVPLGLTANSDTH-LLOGSLTL-- 141
DB 406 LILKQVIFEDDTAIYCCQASNKKGITLANTNVVIELPQILLENENTYTFVGGOKALLEC 465
QY 142 -TLSP-----PGSSPVQCRSPRKNIOGKTLVSQLELDGSGTCTYLOKQKYE 194
DB 466 ETRGSPKRVNTWESSISILLADPRVNLITNG-GLEIANVSHDDSGITCLVQSNISVN 524
QY 195 FKIDI-----VLAFOKASSIVYKKEGOVFSFPLAFVEKLTGSGELMWAQER--ASS 247
DB 525 AVEVLNRTVILSPQALRL--QPKTAIFTC-LYVTDPKLSPLQGRKNDQKIFPSH 580
QY 248 SSKMTTFPLKNEVSVKRVTPKQMGKPLHLTLPLAPQVYAGSGTLTALAKTGK 307
DB 561 SDKKYTFDQGLIIS-----NVEPGDEGYTCQITTKLDMVASSLTLLTC-DRPPP 631
QY 308 LHQEVNLYVMRATQLOKNLTCEWGP---TSPKMLSLKLENKAKVSKREK----- 356
DB 632 VHLQVTNAKHRYVTLN-----WTPGDNNSPILLEYVEFDQMKENGHELRVAAD 684
QY 357 -----PVW 359
DB 685 KGVNLPW 693
RESULT 61
ID IG44 HAEIN STANDARD; PRT; 1849 AA.
AC P45386;
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1995 (rel. 32, Last sequence update)
DE 28-FEB-2003 (rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (GA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxId=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WHI HK61;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases."
RL J. Bacteriol. 174:2913-2921(1992).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A

```

```

CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: Belongs to peptidase family 56.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M87491; AAA24968.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR000710; Peptidase S6.
DR InterPro; IPR004899; Peptactin.
DR Pfam; PF03797; Autotransporter; 1.
DR Pfam; PF03395; IGA1; 1.
DR PRINTS; PR00921; Peptactin; 1.
DR TIGRFAMs; TIGR01414; autotrans bar1; 1.
DR HydroLase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT PROEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
FT ACT SITE 299 299 PROBABLE.
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;
Query Match 4.9%; Score 113.5; DB 1; Length 1849;
Best Local Similarity 19.1%; Pred. No. 9.1; Mismatches 135; Indels 179; Gaps 19;
Matches 88; Conservative 58;
QY 20 PAATQGNKVL-----GKKDVELTCTASQKSIQFHKN 55
DB 735 PHFTENNEVVEDDMINRNFATMTNVTGNASLYSGRVANITSTISNNAQVHIGKT 794
QY 56 SNOIKILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIKNLKIEDSTYICEVEDQK 115
DB 795 GDTVCVRSDYTGVTCHNSNSEKA----- 819
QY 116 EEVQVLVGLT-ANSDTHLLOGSLTLTLSPPGSSPVQCRSPRK-NIOGG-KTLSVS 172
DB 820 -----LNSFNPTNLGNVNLTENASFTL-----GKNLPGTIQSIQTS 857
QY 173 QLELDQSGTWTCTYLOKQKVEFKIDIVLAFQKASSIVYKKEGOVFSFPLAFVEKL 232
DB 858 QVNLKNSHMLTNGSNVNOALNTNGHILHAQDANKVTTYN-----TLTVNSL 907
QY 233 TSGELMWQAERASSSKSWTF-DLKNEVSVKRVTPKQMGKPLHLTLPLAPQY 291
DB 908 SONGSFY-----WVDFTNKSKVNVVNK----- 931
QY 292 AGSGNLTALAKTGK-LHQEVNLY-VMRATQLOKNLTCEWGPSPKMLSLKENKA 349
DB 932 SATGNFTLQVADKTEGEPHNETLFDASNAF--RNNEVLTANGSVRGAWKYLKRVNG 989
QY 350 K-----VSRKRPVWTLN-----PEAGMOCILSDSGQVLLLESNIKVLPTWS 391
DB 990 KYDLNPEVSKRNQVDTNTITTPNDIOADPSA-----QSNNEIARVE 1034
QY 392 TPVHPRASALPAPPTGSLPDPQTASALPDE-PASALPA 430
DB 1035 TPVPP-----PAPATESAI-----ASEQPTRRPAPETAQPA 1064

```

```

RESULT 63
ID      VGR1_MOUSE      STANDARD:      PRT;      1333 AA.
AC      P35969; OS05094; Q61517;
DT      01-JUN-1994 (Rel. 29, Created)
DT      01-JUN-1994 (Rel. 29, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Vascular endothelial growth factor receptor 1 precursor (EC 2.7.1.112)
DE      (VEGFR-1) (Tyrosine-protein kinase receptor FLT (FLT-1) (Embryonic
DE      receptor kinase 2).
GN      FLT1 OR FLT OR EMRK2.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxId=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BALB/c; TISSUE=Neonatal brain, and Placenta;
RX      MEDLINE=93330572; PubMed=8393164;
RA      Finermy H., Kelleher K., Morris G.E., Bean K., Merberg D.M.,
RA      Kriz R., Morris J.C., Sookdeo H., Turner K.J., Wood C.R.;
RT      "Molecular cloning of murine FLT and FLT1."
RL      Oncogene 8:2293-2298(1993).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=94481281; PubMed=8134130;
RA      Choi K., Wall C., Hanratty R., Keller G.;
RT      "Isolation of a gene encoding a novel receptor tyrosine kinase from
RT      differentiated embryonic stem cells."
RL      Oncogene 9:1261-1266(1994).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6; TISSUE=Lung;
RX      MEDLINE=98201626; PubMed=9524283;
RA      Kondo K., Hiratsuka S., Subbalaikshmi E., Matsunishi H., Shibuya M.;
RT      "Genomic organization of the flt-1 gene encoding for vascular
RT      endothelial growth factor (VEGF) receptor-1 suggests an intimate
RT      evolutionary relationship between the 7-Ig and the 5-Ig tyrosine
RT      kinase receptors."
RL      Gene 208:297-305(1998).
CC      -1- FUNCTION: RECEPTOR FOR VEGF AND PLGF. HAS A TYROSINE-PROTEIN
CC      KINASE ACTIVITY. THE VEGF-KINASE LIGAND/RECEPTOR SIGNALING SYSTEM
CC      PLAYS A KEY ROLE IN VASCULAR DEVELOPMENT AND REGULATION OF
CC      VASCULAR PERMEABILITY.
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- SIMILARITY: Belongs to the Tyr family of protein kinases.
CC      -1- SIMILARITY: Belongs to the tyrosine receptor subfamily.
CC      -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC      CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isd-sb.ch/announce/
CC      or send an email to license@isd-sb.ch).
CC      CC
CC      EMBL: L07297; AAA40078.1; -
CC      EMBL: X78668; CA55311.1; -
CC      EMBL: D88689; BAA24498.1; -.
CC      PIR: I78875; I78875.
CC      DR      HSSP: P17948; IGSV.
CC      DR      MGD: MGI:95558; Flt1.
CC      DR      InterPro: IPR007110; Ig-like.
CC      DR      InterPro: IPR003598; Ig_C2.
CC      DR      InterPro: IPR000719; Prot_Kinase.
CC      DR      InterPro: IPR001824; ReceptTykInaIII.

```

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53	Q54	Q55	Q56	Q57	Q58	Q59	Q60	Q61	Q62	Q63	Q64	Q65	Q66	Q67	Q68	Q69	Q70	Q71	Q72	Q73	Q74	Q75	Q76	Q77	Q78	Q79	Q80	Q81	Q82	Q83	Q84	Q85	Q86	Q87	Q88	Q89	Q90	Q91	Q92	Q93	Q94	Q95	Q96	Q97	Q98	Q99	Q100	Q101	Q102	Q103	Q104	Q105	Q106	Q107	Q108	Q109	Q110	Q111	Q112	Q113	Q114	Q115	Q116	Q117	Q118	Q119	Q120	Q121	Q122	Q123	Q124	Q125	Q126	Q127	Q128	Q129	Q130	Q131	Q132	Q133	Q134	Q135	Q136	Q137	Q138	Q139	Q140	Q141	Q142	Q143	Q144	Q145	Q146	Q147	Q148	Q149	Q150	Q151	Q152	Q153	Q154	Q155	Q156	Q157	Q158	Q159	Q160	Q161	Q162	Q163	Q164	Q165	Q166	Q167	Q168	Q169	Q170	Q171	Q172	Q173	Q174	Q175	Q176	Q177	Q178	Q179	Q180	Q181	Q182	Q183	Q184	Q185	Q186	Q187	Q188	Q189	Q190	Q191	Q192	Q193	Q194	Q195	Q196	Q197	Q198	Q199	Q200	Q201	Q202	Q203	Q204	Q205	Q206	Q207	Q208	Q209	Q210	Q211	Q212	Q213	Q214	Q215	Q216	Q217	Q218	Q219	Q220	Q221	Q222	Q223	Q224	Q225	Q226	Q227	Q228	Q229	Q230	Q231	Q232	Q233	Q234	Q235	Q236	Q237	Q238	Q239	Q240	Q241	Q242	Q243	Q244	Q245	Q246	Q247	Q248	Q249	Q250	Q251	Q252	Q253	Q254	Q255	Q256	Q257	Q258	Q259	Q260	Q261	Q262	Q263	Q264	Q265	Q266	Q267	Q268	Q269	Q270	Q271	Q272	Q273	Q274	Q275	Q276	Q277	Q278	Q279	Q280	Q281	Q282	Q283	Q284	Q285	Q286	Q287	Q288	Q289	Q290	Q291	Q292	Q293	Q294	Q295	Q296	Q297	Q298	Q299	Q300	Q301	Q302	Q303	Q304	Q305	Q306	Q307	Q308	Q309	Q310	Q311	Q312	Q313	Q314	Q315	Q316	Q317	Q318	Q319	Q320	Q321	Q322	Q323	Q324	Q325	Q326	Q327	Q328	Q329	Q330	Q331	Q332	Q333	Q334	Q335	Q336	Q337	Q338	Q339	Q340	Q341	Q342	Q343	Q344	Q345	Q346	Q347	Q348	Q349	Q350	Q351	Q352	Q353	Q354	Q355	Q356	Q357	Q358	Q359	Q360	Q361	Q362	Q363	Q364	Q365	Q366	Q367	Q368	Q369	Q370	Q371	Q372	Q373	Q374	Q375	Q376	Q377	Q378	Q379	Q380	Q381	Q382	Q383	Q384	Q385	Q386	Q387	Q388	Q389	Q390	Q391	Q392	Q393	Q394	Q395	Q396	Q397	Q398	Q399	Q400	Q401	Q402	Q403	Q404	Q405	Q406	Q407	Q408	Q409	Q410	Q411	Q412	Q413	Q414	Q415	Q416	Q417	Q418	Q419	Q420	Q421	Q422	Q423	Q424	Q425	Q426	Q427	Q428	Q429	Q430	Q431	Q432	Q433	Q434	Q435	Q436	Q437	Q438	Q439	Q440	Q441	Q442	Q443	Q444	Q445	Q446	Q447	Q448	Q449	Q450	Q451	Q452	Q453	Q454	Q455	Q456	Q457	Q458	Q459	Q460	Q461	Q462	Q463	Q464	Q465	Q466</
----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	--------

```

DB 570 EGEDKLKSCVANKFLYRDTM---ILLRTVNNRTHMHSISKQMATQDYSITL----- 620
QY 93 LIINKLKIEDSDTYICEV-----EDQKEVOLVFGLTANSDPHLLQSGSLTTLTLESP 146
DB 621 NLIVINKVSLSDSGTACRARNITYGEDIIRKTEVLVRSEA---PHLLQNTS---DYEVS 674
QY 147 PGSSPVQCRSPRG-----KN---IQ-----GGKTLVSQLELDQSGTWCT 185
DB 675 ISGSTTLDDCA-KGVPAPOITWFKNNHKIQOEPGIIIGGNTSLIERTEDEGVYRCR 733
QY 186 VLQNKQKVE-----FKIDIVVLAFOKASSIVY-----KKEGEQVEF 221
DB 734 ATNQGAVERSAAYLTVGQTSDKSNLELTCTCVAATLFMLLTFLTRKLRSSSEVYT 793
QY 222 SF-----PLAFTVETKLTGSGELM-WQAEPRASSSSSM-----ITFDLKN- 258
DB 794 DVLSTIMPDDEVLEQOCERLPYDASKFEAREKRLGSLRGAFGVQASAFGIKKS 853
QY 259 ---KEVSVKRTQDPDLQWKKLPLHLTPQALPOVAGSGNTLLALEAKTGKHQEVNLY 315
DB 854 PTCRTVAAYKMLEGATASEYKALMTBLKTLTHGHNLVNLGACTKQGGPL-----MV 908
QY 316 VWRATQLOKNTLCEVWPTSPKMLSLKLENKAVSKREKEVWYNLPAGMOCILSDS 375
DB 909 IVE-----YCK-YGNLSNYL-----KSKRDLFCINKDALHMEIKKES 945
QY 376 GGVLESNIKVLPTWSTPHRPSALPAP-TGSLALPDPQTS 417
DB 946 LEFGLEQCGK-----PRLDSVSSSVTSSSFEDRVS 978

```

RESULT 63

LACH_DROME STANDARD; PRT; 359 AA.

```

ID LACH_DROME STANDARD; PRT; 359 AA.
AC 024372;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lachesin precursor.
GN LAC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=94038693; PubMed=8223276;
RA Karlstrom R.O., Wilder L.P., Baetiani M.J.;
RT "Lachesin: an immunoglobulin superfamily protein whose expression
correlates with neurogenesis in grasshopper embryos.";
RL Development 118:509-522(1993).
CC -1- FUNCTION: MAY PLAY A ROLE IN EARLY NEURONAL DIFFERENTIATION AND
AXON OUTGROWTH.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
similarity).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ON DIFFERENTIATING NEURONAL CELLS
FROM THE ONSET OF NEUROGENESIS IN BOTH THE CENTRAL AND PERIPHERAL
NERVOUS SYSTEMS.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L13255; AAC37184.1; -.
CC HSSP; P56276; TTKL.
CC FlyBase; Fggn0010238; Lac.

```

```

DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00408; IgC2; 2.
DR PROSITE; PS00835; IG_LIKE; 3.
KM Immunoglobulin domain; Cell adhesion; glycoprotein; GPI-anchor;
KW Repeat; Signal; Lipoprotein.
FT SIGNAL 1 25
FT CHAIN 26 336
FT PROPEP 337 359
FT DOMAIN 29 130
FT DOMAIN 135 221
FT DOMAIN 226 317
FT DISULFID 50 113
FT DISULFID 157 204
FT DISULFID 247 303
FT CARBOHYD 92 92
FT CARBOHYD 140 140
FT LIPID 336 336
SQ SEQUENCE 359 AA; 39998 MW; DIF513E2B9D8686 CRC64;

```

Query Match 4.8%; Score 112.5; DB 1; Length 359;
 Best Local Similarity 20.1%; Pred. No. 1.3;
 Matches 46; Conservative 39; Mismatches 109; Indels 35; Gaps 4;

```

QY 34 GDTVELCTASOKSIQPHMKNSQIKILNGSGFLTGPBKNDRADRSRLMDQGNFP 93
DB 43 GGVTFPDCSVQYAEVYNVLTKTSDPVLSTGSLTVIKDSRFSRLDPNNS---TYK 97
QY 94 LIINKLKIEDSDTYICEV-----VEDQKEVOLVFGLTANSDPHLLQSGSLTTLTLESP 148
DB 98 LQIKDIOETDAGTTCQVIVSTVHKVSEVLTSVRPRVID-----NTQSVVASEG 150
QY 149 SSPSVQCRS-----PRGNIOGKTLVSQLELDQSGTWCTVLQNO 190
DB 151 SEVQMECVASGYPTPTITWRRENNALPTDSATVYGNLTIRIKSVKEDRGTYCVADNGV 210
QY 191 KKEVKIDIVVLAFOKASSIYKKKEGEVSESPFLAFTVEKLTGSGELM 239
DB 211 SKGDRNINVEVEFAPVITVPRPLQALQYDMDECHIEAYPPPAIVW 259

```

RESULT 64

C166_CHICK STANDARD; PRT; 588 AA.

```

ID C166_CHICK STANDARD; PRT; 588 AA.
AC P42292;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE CD166 antigen precursor (SCL glycoprotein) (BEN glycoprotein) (DM-
GRASP protein) (UC7 protein).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A. AND SEQUENCE OF 34-53.
RX TISSUE=Embryo;
RX MEDLINE=92030150; PubMed=1931049;
RA Tanaka H., Matsui T., Agata A., Tomura M., Kubota I.,
RA McFarland K.C., Kohr B., Lee A., Phillips H.S., Shelton D.L.;
RT "Molecular cloning and expression of a novel adhesion molecule, SCL.";
RL Neuron 7:535-545(1991).
CC [2]
CC SEQUENCE FROM N.A.
RX MEDLINE=91337449; PubMed=1873027;
RA Burns F.R., von Kamen S., Guy L., Raper J.A., Kamholz J.,
RA Chang S.;
RT "DM-GRASP, a novel immunoglobulin superfamily axonal surface protein
that supports neurite extension.";
RL Neuron 7:209-220(1991).
CC [3]

```

```

PC SEQUENCE FROM N.A. AND SEQUENCE OF 34-52; 87-93; 504-512 AND 569-582.
RP TISSUE-Bursa of fabricius;
RX MEDLINE=92302224; PubMed=1608932;
RA Pourquie O., Corbel C., le Caer J.-P., Rossier J., le Douarin N.M.;
RT "BNR, a surface glycoprotein of the immunoglobulin superfamily, is
RT expressed in a variety of developing systems.";
RN Proc. Natl. Acad. Sci. U.S.A. 89:5261-5265(1992).
[4]
RP POSSIBLE FUNCTION.
RX MEDLINE=92211411; PubMed=113497;
RA Pourquie O., Hallonet M.E.R., le Douarin N.M.;
RT "Association of BEN glycoprotein expression with climbing fiber
RT axonogenesis in the avian cerebellum.";
RL Neurosci. 12:1548-1557(1992).
CC -1- FUNCTION: HOMOPHILIC ADHESION MOLECULE ASSOCIATED WITH CLIMBING
CC FIBER AXONOGESIS. SUPPORTS NEURITE EXTENSION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN A RESTRICTED POPULATION OF AXONS.
CC WITHIN THE SPINAL CORD IT IS LOCALIZED TO AXONS IN THE DORSAL
CC PUNICULUS, MIDLINE FLOOR PLATE CELLS, AND MOTONEURONS. FOUND IN
CC EPITHELIAL, CENTRAL NERVOUS SYSTEM, AND HEMOPOIETIC CELLS. EACH
CC DEVELOPING SYSTEM CARRIES DIFFERENTLY GLYCOSYLATED MOLECULAR FORMS
CC OF BEN.
CC -1- DEVELOPMENTAL STAGE: WIDELY EXPRESSED DURING EMBRYONIC
CC DEVELOPMENT.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like V-type domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; S63276; AAB20170.1; -.
DR EMBL; M76678; AAA48602.1; -.
DR EMBL; X64301; CAA45579.1; -.
DR PIR; A45254; A45254.
DR PIR; JH0464; JH0464.
DR PIR; JH0506; JH0506.
DR HSSP; Q13740; 1K7C.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00409; Ig; 3.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; FALSE_NEG.
KW Cell adhesion, Immunoglobulin domain, Glycoprotein, Transmembrane;
KW Repeat, Signal.
FT CHAIN 1 33
FT DOMAIN 34 588 CD166 ANTIGEN.
FT TRANSMEM 53 532 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 533 553 POTENTIAL.
FT DOMAIN 554 588 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 42 126 IG-LIKE V-TYPE 1.
FT DOMAIN 131 240 IG-LIKE V-TYPE 2.
FT DOMAIN 251 333 IG-LIKE C2-TYPE 1.
FT DOMAIN 338 414 IG-LIKE C2-TYPE 2.
FT DOMAIN 421 501 IG-LIKE C2-TYPE 3.
FT DISULFID 49 119 POTENTIAL.
FT DISULFID 163 226 POTENTIAL.
FT DISULFID 276 319 POTENTIAL.
FT DISULFID 359 397 POTENTIAL.
FT DISULFID 440 490 POTENTIAL.
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

Query Match	Best Local Similarity	4.8%; Score 112; DB 1; Length 588;
Matches 91; Conservative 66; Mismatches 175; Indels 140; Gaps 19;		
FT CARBOHYD	462	462
FT CARBOHYD	485	485
FT CARBOHYD	504	504
FT CONFLICT	1	10
FT CONFLICT	25	25
FT CONFLICT	112	113
FT CONFLICT	329	329
FT CONFLICT	401	402
FT CONFLICT	588 AA;	65726 MW; 2428612D0164531E CRC64;
SO SEQUENCE		
Query Match	4.8%; Score 112; DB 1; Length 588;	
Best Local Similarity	19.2%; Pred. No. 2.6;	
Matches 91; Conservative 66; Mismatches 175; Indels 140; Gaps 19;		
QY 8 RHLLVLVQLALLPRAATGNKRVLGKKQDLYVELTCTASQKKSIOF-HWK---NSNQIKIL 62		
DB 16 RPLCLLALLCWPALGLYTVNAVAVDITIMPCRLLEVPDGLMGKKYKEMPNSPVEFA 75		
QY 63 GNQGSFLITKPSKLANDADRRSLMDQGNPLIILKIKIDSDPYICEV--EQGKEVOL 120		
DB 76 FRSS---TKKNVQVDVDPDYKDRLSISENTYLLISKARISDEKRFVCMVLTEDEVSEPT 132		
QY 121 LVFGLTANSDPTLLQGGSLITLLESPPGSSPSVOCSPRG-----KQ-----IQGKTL 169		
DB 133 VVKFKPSPQSPQELIHOQDPLETEKRLKLGECVYRDSIPEGNVYMKGRVLQPVBEVYVI 192		
QY 170 SVSQGLEQDQSG-----TWCTTVL---QNKQVEF----- 195		
DB 193 NLRKVENRSGLFTMTSSLQYMPTEKDAANKFTCIYVHYGSPGQKTIQSEPVVFDVHYPT 252		
QY 196 -KIDIVVLAVQKASLYKKKEGVEYSPFLAFVETLTSGE-----LW---QAKRA 245		
DB 253 EKVIRIRLS--QSSSTI---KEGNAVTL-----KSGNGNPPQGEFLFYIPGEETG 298		
QY 246 SSSKSMITPDKKNEVSVRKVTQDPK-----LQMGKRLPLH 281		
DB 299 RSSSTYMTVDYRRNATGEYKCSLIDKSMNATITTYHYDLQLTSPGEVTKQIGEALPVS 358		
QY 282 LTLPLPALP-----QYAGSGNL---TLALEAKTGLKHOENVLV- 316		
DB 359 CTISSSNATVFWIKDTRAKTSPSPSLQYDQAGNYICETTLIQEYVGLKKRKTLLKIVE 418		
QY 317 -----MRATQLQKLTCEYWGTSFKMLSLKLENKEAKVSKREKPVVY 360		
DB 419 GKPIKMTKTKTNTKMSKTIYCHVEG--FPKPAVQMTVTGSGSLINKTEETKYV 470		
RESULT 65		
IL1S_CERAE	STANDARD;	PRT; 393 AA.
IL1S_CERAE	Q29612;	
DT 15-JUL-1999 (Rel. 38, Created)		
DT 15-JUL-1999 (Rel. 38, Last sequence update)		
DT 10-OCT-2003 (Rel. 42, Last annotation update)		
DE Interleukin-1 receptor, Type II precursor (IL-1R-2) (IL-1R-beta).		
GN IL1R2 OR IL1RB.		
OS Cercopithecus aethiops (Green monkey) (Grivet).		
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;		
OC Cercopithecinae; Cercopithecus.		
OX NCBI_TaxID=9534;		
RN RN	SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT), AND SEQUENCE OF 125-145.	
RP RP	MEDLINE=96355446; Pubmed=8702856;	
RA RA	Liu C., Hart R.P., Liu X.J., Cleverger W., Maki R.A., Souza E.B.;	
RT RT	"Cloning and characterization of an alternatively processed human	
RL RL	type II interleukin-1 receptor mRNA."	
CC CC	J. Biol. Chem. 271:20965-20972 (1996).	
CC CC	-1 FUNCTION: RECEPTOR FOR INTERLEUKIN-1 ALPHA (IL-1A), BETA (IL-1B),	
CC CC	AND INTERLEUKIN-1 RECEPTOR ANTAGONIST PROTEIN (IL-1RA).	
CC CC	-1 SUBCELLULAR LOCATION: THE LONG ISOFORM IS A TYPE I MEMBRANE	
CC CC	PROTEIN WHILE THE SHORT ISOFORM IS SOLUBLE.	
CC CC	-1 ALTERNATIVE PRODUCTS:	

```
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q29612-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q29612-2; Sequence=VSP_002663;
CC -1- SIMILARITY: Belongs to the interleukin-1 receptor family.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U64092; AAB05876.1; -.
CC EMBL; U64093; AAB05877.1; -.
CC HSRP; P14778; IITB.
CC InterPro; IPR007110; IG-like.
CC InterPro; IPR003599; IG.
CC InterPro; IPR004074; IL1_receptor/IL1.
CC InterPro; IPR004077; IL1_receptor/IL1.
CC Pfam; PF00047; Ig; 3.
CC PRINTS; PR01539; INTERLEUKINR2.
CC PRINTS; PR01536; INTERLEUKINR12F.
CC SMART; SM00409; IG; 3.
CC PROSITE; PS50835; IG_LIKE; 3.
CC Immunoglobulin domain; Receptor; Glycoprotein; Transmembrane; Signal;
CC Repeat; Alternative splicing.
CC FT CHAIN 1 133 INTERLEUKIN-1 RECEPTOR, TYPE II.
CC FT SIGNAL 1 393 EXTRACELLULAR (POTENTIAL).
CC FT DOMAIN 14 348 POTENTIAL.
CC FT TRANSMEM 368 367 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 29 120 IG-LIKE C2-TYPE 1.
CC FT DOMAIN 134 221 IG-LIKE C2-TYPE 2.
CC FT DOMAIN 237 342 IG-LIKE C2-TYPE 3.
CC FT DISULFID 50 108 BY SIMILARITY.
CC FT DISULFID 152 207 BY SIMILARITY.
CC FT DISULFID 258 326 BY SIMILARITY.
CC FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT VARSPLIC 297 393 Missing (in isoform Short).
CC FT FTId=VSP_002663.
CC SQ SEQUENCE 393 AA; 44778 MW; DAD746C5DE59A8 CRC64;

Query Match 4.8%; Score 111.5; DB 1; Length 393;
Best Local Similarity 20.3%; Pred. No. 1.7;
Matches 70; Conservative 52; Mismatches 120; Indels 103; Gaps 14;

Oy 7 FRHLVLVLA---LLPATOG---NKVVLGK-----KGDYVELTCT-----ASQ 45
Db 2 FRLLVAVGVSAFTQPAHAHTGAASCPYRHHYKREFRLEBEPALACPOVPYVLAHSV 61
Oy 46 KKSIOFHNKSNQIKILGNQGSFLTKGPSKLNDRADSRSLMDQGNFLLIKNLTIEDSD 105
Db 62 SPHINLTVHKNDASAMV-----PGEETR-----WMAQGALMLPALQ-EDSG 104
Oy 106 TYICEVED---QKEQVLLVFGLTANSPTHLLOQSILTL-----TLSEPP 147
Db 105 TYICTTRNASYCDKYSIELRPFENTDASLPFISYDQIITLSTFGVLVCPDLREFTRDXTD 164
Oy 148 GSPSPVOCSPRGK-----NIQGGKTLVSQLEODSGTMTCTV-----186
Db 165 GKIQWYKQFLPDKNKEKELSVRGTHLLVHVALIEDAGYRVCVLTFAHEGQYITRNI 224
Oy 187 -LQNGKVEFKIDIVLAFQKASIVYKKEGEQVFSFPLAFTVEKLTGSGE-----LMW 240
Db 225 ELRIKKKEETIPVILISPLKTIASLSGR-----LTIPKAVPFGTGPPLTTLMLM 274
```

```
Oy 241 QAERASSKSWITFDLKNKEVSVKRVODPKLQWKKLPLHLTP 285
Db 275 TANDTHESAY-----PGGRVTEGPEQREISENNENTIEVP 309

RESULT 66
SCA4_RICFE STANDARD; PRT; 981 AA.
ID SCA4_RICFE
AC Q9AJ37;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia felis (Rickettsia azadi).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=42862;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing the sequence of
RT gene D coding for an intracytoplasmic protein.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF196973; AAK1304.1; -.
CC Antigen.
CC FT NON_TER 1 1
CC FT NON_TER 981 981
CC SQ SEQUENCE 981 AA; 107514 MW; 7F18F421E2C262E1 CRC64;

Query Match 4.8%; Score 111.5; DB 1; Length 981;
Best Local Similarity 24.4%; Pred. No. 5.5;
Matches 88; Conservative 48; Mismatches 150; Indels 75; Gaps 19;

Oy 110 EVEDQKEVQLVGLTANSD--THLQGSILTLLES--PRGSSPSYQCRSPRGNTQGG 166
Db 103 EIEBKQKTLISAFQGNPANREFIDKALENPELKKLSIEIAGYKNVLSTVSANGYGG 162
Oy 167 -----KTLVSQLEL---QDSGTWTCTYLVQNGKYEFKIDIVLAFQKASIVYKKE 215
Db 163 FKPYQWENQISASPLRATVVNNDAGDELCTL--NETTYKTK-----PFTVAKQD 209
Oy 216 GEQV-----EFSPPLAFTVEKLTGSGELMQAERASSKSWITFDLKNKEVSVKRVQD 269
Db 210 GTQVQINSYREIDPPI--KLDKADGSHLNSVALKADSTK-----PSKQKAVYFTAHY 260
Oy 270 PKLQWKKLPLHLTPQALPOYAGSGNLTALAEKTKLHDEQVNLVVMARATOLQKNLTCE 329
Db 261 EEGPNQKQQLKEISPKRL-KFADGPDPAVYIEHGGLY---TLAVTRGK--YKEMWRE 314
Oy 330 VWGFTSPKLMSTKLKNEKAVYSKR--EKPYWVLNPEAGMWQCLSDSGVLLSNTIKVL 387
Db 315 VELNQGSVDLSQRTIADLTIVQGRSQETPQPIITPQOE-----LKSISIT- 360
Oy 388 PTWSPVYPRASA-LPAAPPTGSAIPDQGTASALDPP-----AASALPAALAVISFLGLG 442
Db 361 PT-TTQVPPITPAPQDLPQETSQMPQOVN-----PMLVATYATLSTSMQDLNATYVAG 414
Oy 443 L 443
Db 415 L 415
```

```

RESULT 67
FPRP_MOUSE
ID_FPRP_MOUSE STANDARD; PRT; 879 AA.
AC Q9W9J1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Prostaglandin F2 receptor negative regulator precursor (Prostaglandin
DE F2-alpha receptor regulatory protein) (Prostaglandin F2-alpha receptor
DE associated protein).
GN PIGRN OR FPRP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP Weng L., Falla N., Van Den Heuvel R., Van Vlaselaer P.,
RA Raymakers J., Lowik C., Meregault J.;
RT "The monoclonal antibodies 18D7/91F2 recognize a receptor regulatory
RT protein on mouse bone marrow stromal cells.";
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: Inhibits the binding of prostaglandin F2-alpha (PGF2-
CC alpha) to its specific FP receptor, by decreasing the receptor
CC number rather than the affinity constant. Functional coupling with
CC the prostaglandin F2-alpha receptor seems to occur (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Endoplasmic
CC reticulum and trans-Golgi network (By similarity).
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, AF152344; AAD38383.1; -.
DR MGD; MGI:1277114; Pctgm.
DR GO; GO:0005515; F:protein binding; ISS.
DR GO; GO:0017148; P:negative regulation of protein biosynthesis; ISS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 6.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG LIKE; 5.
KW Signal, Immunoglobulin domain, Transmembrane, Glycoprotein, Repeat.
FT SIGNAL 1 21
FT CHAIN 22 879
FT
FT REGULATOR.
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM POTENTIAL.
FT DOMAIN 833 853
FT DOMAIN 854 879
FT DOMAIN 22 137
FT DOMAIN 149 263
FT DOMAIN 276 389
FT DOMAIN 406 536
FT DOMAIN 544 662
FT DOMAIN 688 813
FT DOMAIN 424 427
FT SITE 89 91
FT SITE 703 705
FT DISULFID 43 119
FT DISULFID 169 247
FT DISULFID 299 373
FT DISULFID 429 515
FT DISULFID 571 655
FT DISULFID 711 793
FT CARBOHYD 44 44

```

```

FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 525 525 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 600 600 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 618 618 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 691 691 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 879 AA; 98707 MW; 0D3D3EE271478D92 CRC64;
Query Match 4.84; Score 110.5; DB 1; Length 879;
Best Local Similarity 19.54; Pred. No.5.6; 203; Indels 125; Gaps 21;
Matches 100; Conservative 86; Mismatches
QY 14 LQALPAATQGNKVVGLKGGDTVELTCTASQ-----KKSIOFMKNSNQIKLG---- 63
Db 279 LQLAV-----PRYSVTEGKDLDSCLNTTRVDVDRVETVMYFKKTDLSLASHML 331
QY 64 ---NQGSFLTKGPKSLNDRADRSRLMDQGNFLIIKLIKIEDSDYICV-----ED 113
Db 332 ARLDRLSLVHSPHALSHVDR-----SYHLVADVSKENSGYLCVALMAPGHR 384
QY 114 QKEEYQLVGLTANSDTHLQGSULTLESPPGS---PSVQCRSPRGKNIQGGKTL 170
Db 385 SMKIVAEANSPGVSVTWLEPEYQVYLNSKVPGEFDDPTLEQCRVIDTKLEAGVRLT 444
QY 171 V-----SOLELDGSTMCTVLQNOKKVEFKIDIVLAFQKA--SSIVYK 213
Db 445 VSMYTRNRNDVVASSELLAVMDGDTLRYGERSK-----QRAQDGEIF 491
QY 214 KEGBQVEFSPPLAFYVEKLTGSGELMWQAEASSKSMV--TFPLKKNKSVKRYTDPKL 272
Db 492 KEHTDTR--FNRIORTTEEDBNYCVVASAWARQNNMSVKSQVFSKPVNI FNASEBSVL 550
QY 273 QMGKKPLHLTLPOALPOVAGSNLTLAEAKTGKLGQEVNLVWVRATOLQKULTCEW 332
Db 551 VVKARQD-----KPFPAAGTFEWTCKVSSKNIKSPYSLITKEKVGDIS---S 598
QY 333 PTPSKMLTSL-----KLEN-----KEAKVSREKRPVWLNPEAGMOCQL- 372
Db 599 PNEFKYIISLDQSVYKLEWMTDASRDVGLKQVDEDERHYMYQQSDADGLYRCMVT 656
QY 373 --SDSGCVL-----LESNIKV---LPTWSTPVHPRASALPAPETGALP--DPQT 415
Db 659 AWPPIGGSLTREAATSNPIEIDFQTSGETFNASVH---SDTPSVYRGDLKLCFIVTV 715
QY 416 ASALPDPPASALPALAVISFLGLGIVACVL 449
Db 716 EGAVLDPDDMAFDVSMFAVHSF---GLDKAPVL 745
RESULT 68
RTN1_RAT
ID_RTNI_RAT STANDARD; PRT; 777 AA.
AC Q64546; Q64547;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Reticulon 1 (Neuroendocrine-specific protein) (S-rex).
GN RTN1 OR NSP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN
RP SEQUENCE FROM N.A. (ISOFORMS RTN1-B AND RTN1-S).
RC STRAIN=Wistar; TISSUE=Brain cortex;
RX MEDLINE=96386034; PubMed=8793864;
RA Baka I.D., Ninkina N.N., Pinon L.G.P., Adu J., Davies A.M.,
RA Georgiev G.P., Buchman V.L.;
RT "intracellular compartmentalization of two differentially spliced s-
RT rex/NSP mRNAs in neurons.";
RL Mol. Cell. Neurosci. 7:289-303 (1996).
CC -1- FUNCTION: May be involved in neuroendocrine secretion or in

```



```

CC membrane trafficking in neuroendocrine cells.
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum membrane (By
CC similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=RTN1-B; Synonyms=S-RexB;
CC IsoId=064548-1; Sequence=Displayed;
CC Name=RTN1-S; Synonyms=S-RexS;
CC IsoId=064548-2; Sequence=VSP_005647, VSP_005648;
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN CENTRAL AND
CC PERIPHERAL NERVOUS SYSTEM OF NEWBORN AND ADULT RATS. LOW LEVELS
CC HAVE BEEN ALSO DETECTED IN HEART, ADRENAL GLAND AND SPLEEN.
CC EXPRESSION OF ISOFORM RTN1-B IS RESTRICTED TO PARTICULAR NEURONAL
CC TYPES.
CC -1- DEVELOPMENTAL STAGE: DETECTED ON EMBRYONIC DAY E10 IN THE
CC HINDBRAIN AND IN E11 IN THE FOREBRAIN. DURING THE NEXT 3 EMBRYONIC
CC DAYS THE LEVELS OF S-REXS INCREASES AND REMAINS STABLE AT E13 IN
CC THE HINDBRAIN AND AT E14 IN THE FOREBRAIN. THE LEVELS OF S-REXB
CC DOES NOT CHANGE AS SIGNIFICANTLY DURING DEVELOPMENT OF THE
CC HINDBRAIN.
CC -1- SIMILARITY: Contains 1 reticulation domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; U17604; AAC53046.1; -
CC DR InterPro; IPR003388; Reticulon.
CC DR Pfam; PF02453; Reticulon; 1.
CC DR PROSITE; PS50845; RETICULON; 1.
CC DR Endoplasmic reticulum; Alternative splicing; Transmembrane.
CC FT TRANSMEM 604 624 POTENTIAL.
CC FT DOMAIN 550 747 POTENTIAL.
CC FT DOMAIN 550 777 RETICULON.
CC FT VARSPLIC 1 569 POLY-LEU.
CC FT VARSPLIC 570 589 Missing (in isoform RTN1-S).
CC FT /FTId=VSP_005647.
CC FT IGPISGDIPLPPLPENNCK -> MQATADSTKMDQWSNW
CC FT KSQ (in isoform RTN1-S).
CC FT /FTId=VSP_005648.
CC FT SEQUENCE 777 AA; 83001 MW; AF7479C5DF2BD0AC CRC64;
CC -----
CC Query Match 4.7%; Score 110; DB 1; Length 777;
CC Best Local Similarity 19.4%; Pred No. 5.1;
CC Matches 77; Conservative 50; Mismatches 142; Indels 128; Gaps 15;

```

```

DB 483 QDSPMKP--GVLDAIRETSSRATEER---APSHQGVPEBDDPILSFTPTVLQGRPEBSS 537
QY 416 ASALPDP-----PAAALPAA 432
DB 538 GDGAPVPEPPKXQQQKPEEAHVSSSQSPAAATPEIGPL 574

RESULT 69
GCC2_MOUSE STANDARD, PRT, 1679 AA.
AC Q8CHG3; Q8BR44; Q8R205; Q9CT45;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE GRP and coiled-coil domain-containing protein 2 (Golgi coiled coil
DE protein GCC185).
GN GCC2 OR KIAA0336.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OK NCBI_taxid=10090;
RN [1]
RP SEQUENCE OF 1-769 FROM N.A.
RC TISSUE=Breast cancer;
RX MEDLINE=22388257; PubMed=12477932;
RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedtin T.B., Toshitsuki S., Carrini P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Heltan E., Keltman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schin J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE OF 1-442 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain; and Embryo;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nishida I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Haegawa Y., Nogami A., Schonbach C., Gotohori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Brad D., Brusic V., Chochia C., Corbani L.E., Cousins S.,
RA Datta E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimwood S., Guecinich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Kongaya A., Kurichin I.V., Lee Y., Lehman R.D., Lyons P.A.,
RA Maggiori D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Nunata K., Okido T., Pavan W.J., Petrea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Waranabe Y., Wells C.,
RA Wilting L.G., Wyszewski B., Yang A., Yangisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Zimin P., Hayatsu N.,
RA Hitozane-Tshikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shitaka T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Iehil Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,

```

RA Yaenishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RT Nature 420:563-573(2002).
 RN [3]
 RN SEQUENCE OF 49-1679 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22579291; PubMed=12693553;
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
 RA Nakajima D., Nagase T., Ohara O., Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
 RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
 RT cDNAs identified by screening of terminal sequences of cDNA clones
 RT randomly sampled from size-fractionated libraries";
 RL DNA Res. 10:35-48(2003).
 CC -1- FUNCTION: Probably involved in maintaining Golgi structure (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; peripheral membrane protein
 CC associated with the trans-Golgi network (By similarity).
 CC -1- DOMAIN: Extended rod-like protein with coiled-coil domains.
 CC -1- SIMILARITY: Belongs to the golgin family.
 CC -1- SIMILARITY: Contains 1 GRIP domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 DR EMBL; BC027339; AA027339.1; ALT_INIT.
 DR EMBL; AK011206; BAB27466.2; -;
 DR EMBL; AK045701; BAC32463.1; -;
 DR EMBL; AB093232; BAC41416.1; -;
 DR PIR; P70649; P70712.
 DR MGD; MGI:1917547; 2600014C01R1K.
 DR InterPro; IPR000237; GRIP domain.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF01465; GRIP; 1.
 DR PROSITE; PS50913; GRIP; 1.
 DR KMW; Golgi stack; Coiled coil.
 FT DOMAIN 31 1613 COILED COIL (POTENTIAL).
 FT DOMAIN 83 501 GLU-RICH.
 FT DOMAIN 1604 1654 GRIP.
 FT CONFLICT 440 440 I--> M (IN REF. 2).
 SQ SEQUENCE 1679 AA; 194443 MW; 6E2F84263E5E0103 CRC64;
 Query Match 4.7%; Score 110; DB 1; Length 1679;
 Best Local Similarity 20.0%; Pred. No. 14;
 Matches 87; Conservative 67; Mismatches 133; Indels 148; Gaps 18;
 Oy 15 QALIPATQGNKVVLGKGDPTVELTCTASOKSISQ--FWKNSNQIKITIGNGSFLTK 71
 Db ELRLTVLRVSEDEKILTRK---ELDAVTSAKKALQDLLEMKONTNKALENQ--TLST 871
 Oy 72 GPSKLNDRADSRSRSLMDQ-----GNPLTIKN-----LK-----IE 102
 Db 872 QVEELISQTLHSRNEVHDEKVLVIEHENIRLLKQRESELQVRAELIILKSLSEKSPSVK 931
 Oy 103 DSDTYICEVEDQKE-----EVLVLFGLTANSDTHTLQGSITLTLESPPG 148
 Db 932 DQLSLVLEKEIKESKDEKSKIKLVAVKAKKELDSNRKQGTRELELS--- 988
 Oy 149 SSPVQCSPPGKXNIQGSKITSVSGLE-LQDSGWTCTVLOKVKVEKIDIVLAPKA 207
 Db 989 -----VASEKDR-----LSASKKEFLQCAESYKSLLE----- 1016
 Oy 208 SSIYKKEGGEVSEFPLATFVEKLTGSGELMWQAEKSSKSMI--TPDLKKKEVSVK 264
 Db 1017 ----YDKOSEQD-----VEK-----ERAHNFERHIEDLTQKLRNSTOCYE 1053

Oy 265 RVTQDPK-----LQMGKGLPLHLTPQALPOYAGSGLTLALNAKTKLHQEVLNV 316
 Db 1054 RLTSNDNEDLARIETTLQANK-----LLEAQILLEVQXAKGVVEKELDAEBLQKEGKIEHV 1109
 Oy 317 MRATOLQKNTLCEVWGTSFKMLSLKLENKEAK-----VSKREKPVWVLPENKGM 368
 Db 1110 STVNELE-----ELQLQFQKEKKQDLQKTMQEBELVYKDAQQTLLNMMEIADY 1156
 Oy 369 QCLSDSGVYLLLESN 383
 Db 1157 ERLMKELNQKLTNKN 1171
 RESULT 70
 UN89 CAEEL STANDARD; PRT; 6632 AA.
 AC 001761; 017362;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
 GN UNC-89 OR C09D1.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RN SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC STRAIN=Briscot N2;
 RX MEDLINE=96180278; PubMed=8603916;
 RA Benlian G.M., Tinley T.L., Tang X., Borodovsky M.;
 RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
 RT transduction domains";
 RT J. Cell Biol. 132:835-848(1996).
 RL [2]
 RL SEQUENCE FROM N.A.
 RC STRAIN=Briscot N2;
 RA Du Z., Le T.T., Wilson R.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN REVISIONS.
 RP Waterston R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Structural component of the muscle M-line. Myofibril
 CC lattice assembly begins with positional cues laid down in the
 CC basement membrane and muscle cell membrane. UNC-89 responds to
 CC these signals, localizes, and then participates in assembling an
 CC M-line.
 CC -1- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
 CC -1- SIMILARITY: Contains 1 DBL-homology (DH) domain.
 CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
 CC -1- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
 CC -1- SIMILARITY: Contains 1 PH domain.
 CC -1- SIMILARITY: Contains 5 RGS domain.
 CC -1- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 DR EMBL; U33058; AA00542.1; -;
 DR EMBL; AF003131; AAB54132.2; -;
 DR PDB; 1FHO; 20-DEC-00.
 DR WormPep; C09D1.1; CE30426.
 DR InterPro; IPR008957; FN III-like.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR007110; Ig_c2.
 DR InterPro; IPR003598; Ig_c2.

DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001849; PH_
DR InterPro; IPR007850; RCSD.
DR InterPro; IPR000219; RhGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; Ig; 47.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF05177; RCSD; 5.
DR Pfam; PF00621; RhGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGC2; 23.
DR SMART; SM00325; RhGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50010; DH_2; 1.
DR PROSITE; PS50835; IG_LIKE; 49.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW Muscle protein, immunoglobulin domain; Repeat; SH3 domain;
3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1882 2067 IG-LIKE C2-TYPE 7.
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.

FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBRONECTIN TYPE-III.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.
FT DISULFID 5946 5998 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT CONFLICT 2137 2137 A -> P (IN REF. 1).
FT CONFLICT 2245 2247 AKA -> PKP (IN REF. 1).
FT CONFLICT 2258 2258 A -> P (IN REF. 1).
FT CONFLICT 2284 2284 E -> G (IN REF. 1).
FT CONFLICT 2297 2297 M -> I (IN REF. 1).
FT CONFLICT 3531 3531 A -> G (IN REF. 1).
FT CONFLICT 3884 3888 DAGET -> KRRRI (IN REF. 1).
FT CONFLICT 3929 3929 A -> V (IN REF. 1).
FT CONFLICT 5145 5145 A -> P (IN REF. 1).
FT CONFLICT 5146 5145 T -> S (IN REF. 1).
FT CONFLICT 5185 5185 G -> A (IN REF. 1).
FT CONFLICT 5189 5185 K -> N (IN REF. 1).
FT CONFLICT 5202 5202 L -> F (IN REF. 1).
FT CONFLICT 5213 5213 F -> L (IN REF. 1).
FT CONFLICT 6178 6178 A -> G (IN REF. 1).
FT CONFLICT 6268 6268 K -> E (IN REF. 1).
SQ SEQUENCE 6632 AA; 731665 MW; 262D3EDD62960B89 CRC64;

Query Match 4.7%; Score 110; DB 1; Length 6632;

Best Local Similarity 18.7%; Pred. No. 83;
Matches 97; Conservative 60; Mismatches 132; Indels 230; Gaps 25;

QY 32 KKGDVELTLCASQKSIQPHM-KNSNOIKILNGGSLTIGKPSKLNDRADRSRLMVG 90
DB 2899 KKGSAVEECVVPDTGVCCKWLDGKEILARIIVOTRTGP-----EG 2943
QY 91 NF-PLIINKLIEDSDTYICEVEDQKEVOLLVFGLTANSDFHLQGSYTLT-LSSPP 147
DB 2944 HITQSLVLDNTPEDAGKTYCIVEN-----TAGDGT-----CENTLVIESLE 2986
QY 148 GSSPS-----VQCR--SPRGK-----NIQG 165
DB 2987 KKSEKAEDEFIVALQDKTTTSEKVLCEKVIYGEKPKVSMIDHNKITIQESITVESVEG 3046
QY 166 GKTLVSQSLQDSTGTCVYLQNGKKVEFDIVLVAFOKASSIVYKKEBEQVFEFPL 225
DB 3047 VERVITTSLSHQCKYTC-----IAENTETGSKTEAF-- 3079
QY 226 AFTVEKLGSSELMWQAEARASSKSWITPDLKNKEVSRYKVTQPKLOMGKKLPLHLTL- 284
DB 3080 -----LVQGE-----APVFTKEIQNKELSI-----GKLVLSGSVK 3111
QY 285 --PQALPOYAGSGNLT-----LALDAKTKLHQEVN---LVVMRATQ----- 321
DB 3112 GSPQGHVDFVFSEFTTKVETKITSSRIAIE-----HDQNTMWRMVISQITKEDIIVSY 3165
QY 322 -----LQKNLTCEVWGP-----TSQPKMLSLKLNK-----EAKVSRRE 355
DB 3166 KAIATNSIGTATSTSKITTKVEAPVFEQGLKKTYSKEKEIKMEVKVGGASPAVDEWFEKDD 3225
QY 356 KPV-----VWLNPEAGMW-----QCLSDSGGVLTLESNITKVLPTWSTPVHPASALPA 403
DB 3226 KPVSDGNHKKKRPETGVFLVLVKQALTTDAGKYTAA-----SNPAGTASSSAA 3277

